UNIVERSITY OF CAPE TOWN

DEPARTMENT OF ASTRONOMY

MSC THESIS

Optimisation of Galaxy Identification Methods on Large Interferometric Surveys

Author:
THEMBA GQAZA

Supervisor:
PROF. R.C. KRAAN-KORTEWEG
PROF. T. JARRETT

A thesis submitted in fulfillment of the requirements
for the degree of Masters in Astrophysics and Space Science.

December 2018

Abstract

THEMBA GQAZA

Optimisation of Galaxy Identification Methods on Large Interferometric Surveys

The astronomical size of spectral data cubes that will result from the SKA pathfinders planned large HI surveys such as LADUMA; Fornax HI survey; DINGO; WALLABY; etc. necessitate fully automated three-dimensional (3D) source finding and parametrization tools. A fraction of the percentage difference in the performance of these automated tools corresponds to a significant number of galaxies being detected or undetected. Failure or success to resolve satellites around big spirals will affect both the low and the high mass end of the HI mass function. As a result, the performance and efficiency of these automated tools are of great importance, especially in the epoch of big data. Here I present the comprehensive comparison of performance between the fully automated source identification and parametrization software: SOFIA, the visual galaxy identification method and the semi-automated galaxy identification method. Each galaxy identification method has been applied to the same $\sim$ 35 gigabytes 3D HI data cube. The data cube results from the blind HI imaging survey conducted using the Westerbork Synthesis Radio Telescope (WSRT). The survey mapped the overdensity corresponding to the Perseus-Pisces Supercluster filament crossing the Zone-of-Avoidance (ZoA), at $(\ell, b) \approx (160^{\circ}, 0.5^{\circ})$.

A total of 211 galaxies detected using the semi-automated method by Ramatsoku et al. [2016]. In this work, I detected 194 galaxies (using the visual identification method) of which 89.7% (174) have cross-matches/counterparts on the galaxy catalogue produced through semi-automated identification method. A total of 130 detections were made using SOFIA of which 89 were also identified by the two other methods. I used the sample of 174 visual detections with semi-automated counterparts as a Testbed to calculate the reliability and completeness achieved by SOFIA. The achieved reliability is $\sim$ 0.68 whereas completeness is $\sim$ 0.51. Further parameter fine-tuning is necessary to have a better handle on all SOFIA parameters and achieve higher reliability and completeness values.

Keywords: techniques: interferometric-methods: data analysis-radio lines: galaxies

Acknowledgements

I would like to thank my supervisor and mentor Prof. R.C. KRAAN-KORTEWEG of the Department of Astronomyat the University of Cape Town (UCT). Whenever I ran into a wall with my thesis or had questions or facing challenges with writing, she was there to guide me in the right direction. Even when I was not coping with pressure, feeling defeated and uninspired, she made it a point to reach out and pull me from the hole then reignite the passion within me. She consistently challenged me to go an extra mile. She allowed this project to be my own work, but steered me in the right the direction whenever she thought I needed it.

I express my gratitude to Prof. T. JARRETT of the University of Cape Townfrom UCT for being my co-supervisor and for funding. His encouragement, insightful comments, challenging questions and passion for Astronomy inspired me greatly. His involvement improved the quality of this project.

I acknowledge Dr. B. FRANK from the Inter-University Institute for Data-intensive Astronomy (IDIA) for the role he played in my thesis. He taught me to always be critical of my work and put emphasis on assessing my results rigorously.

I must thank Dr. M. RAMATSOKU from the Osservatorio Astronomico Cagliari (OAC) for producing both the data cube and Semi-automated source catalogue used in this project. Dr. K. SAID from the Australian National University gave me great input on deriving HI parameters and fitting noise baselines using various tools available to astronomers.

I would also like to thank the experts who help put together the Visual source catalogue presented in this thesis: Dr. A.C. SCHRÖDER from the South African Astronomical Observatory (SAAO) and Prof. P.A. HENNING from the University of New Mexico (UNM). Without their participation and input, the Visual source catalogue could not have been successfully derived. I must also thank Dr. E.C. ELSON at the Department of Physics and Astronomy at the University of Western Cape (UWC) for the mocked spectral data cubes he created for this project. I am also grateful for the great interactions and input I had with Dr. P. SERRA from OAC with regards to SOFIA.

Finally, I must express my profound gratitude to my family and friends, especially ALINA SONJICA and GCOBISA NGEDLE for the continuous support and encouragement throughout my years of study. This achievement would not have been possible without them.

Contents

List of Figures

List of Tables

List of Symbols

ℓ	deg	Galactic latitude
b	deg	Galactic longitude
z	$-$	Redshift
ν	Hz	Frequency
D	Mpc	Cosmological distance
V_{hel}	$km\,s^{-1}$	Heliocentric velocity
W_{20}	$km\,s^{-1}$	Linewidth at 20% of the peak flux
W_{50}	$km\,s^{-1}$	Linewidth at 50% of the peak flux
S_{peak}	$Jy\,beam^{-1}$	Peak flux
W_{int}	$Jy\,km\,s^{-1}$	Integrated flux
SNR	$-$	Signal-to-noise ratio
M_{HI}	$M_\odot$	HI mass

$\mathfrak{C}$	Completeness	
$\mathfrak{R}$	Reliability	
$\mathcal{L}_{3\sigma,100}$	———	3σ detection threshold while assuming $W_{50} = 100\,km\,s^{-1}$
$\mathcal{L}_{3\sigma,200}$	———	3σ detection threshold while assuming $W_{50} = 200\,km\,s^{-1}$

Constants

Hubble constant	H_0	$70\ km\,s^{-1}\ Mpc^{-1}$
Speed of light	c	$299\,792\,458\ m\,s^{-1}$
HI line frequency	ν_{HI}	1 420.405751 MHz

Definitions

Fine-tune $-$ tweak or adjust control parameters to achieve more accurate results.
Testbed $-$ a catalogue that contains galaxies identified by both the Visual and Semi-automated source finding method.

Chapter 1

Introduction

The precursors to the Square Kilometre Array (SKA: Carilli and Rawlings [2004]) like the South Africa Karoo Array Telescope (MeerKAT: Booth et al. [2009]), the Australian SKA Pathfinder (ASKAP: Johnston et al. [2008]) and the Westerbork Synthesis Radio Telescope (WSRT) with Apertif (APERTIF: Verheijen et al. [2008]) will conduct large-scale extragalactic HI surveys which will produce unprecedented amounts of data in short periods. The expected high influx of data requires advanced automated pipelines to deal with the challenges of working with big datasets such as signal capture; data mining; storage; data processing speed and visualization. As a result research groups such as the Inter-University Institute for Data Intensive Astronomy based in South Africa[1] (IDIA) and the Australian Pawsey Supercomputing Centre[2] have large teams of international researchers focusing on building pipelines to tackle some of these challenges.

One of the greatest challenges involves differentiating genuine HI line emission from the unwanted signals in the data. Most of the perturbing signals come from continuum emission and radio frequency interference (RFI). The signal from these contributors is generally much brighter than the HI line emission signal present in the data. Sophisticated software for removing continuum and flagging RFI signals exist but are not perfect. Hence final data products might show none-flat baselines and unidentified RFIs (during the flagging step) might reduce the image quality, which will reduce the probability of identifying faint galaxies.

The most advanced automated source finding pipelines use matched filtering to identify HI line emission from the data. However unmitigated RFIs and baseline residuals left in the data possess significant signal, hence they result in a high number of false detections. Correcting for this bias can be a tedious task and involves a series of parameter steps. In addition to employment of automated source finding such as MULTIFIND (Kilborn [2001]), recent large HI surveys based on single dish observations (e.g. the HI Parkes All-Sky Survey (HIPASS: Meyer et al. [2004]) and the Arecibo Legacy Fast ALFA Survey (ALFALFA: Giovanelli et al. [2005])) involved a significant amount of visual examination when it came to classifying detections as either genuine or false positive and parametrisation. The epoch of big data requires fully automated source finding and parametrisation

[1] http://idia.ac.za/about-us
[2] https://www.pawsey.org.au/research/

software.

Flöer et al. [2014] developed a fully automated source finding, parametrisation and classification for the extragalactic Effelsberg-Bonn HI Survey. The pipeline employs machine learning techniques such as an artificial neural network (ANN) to classify the source as genuine of false positive and it arrives at that decision by using derived HI parameters. However, this software has only been applied to a single dish HI data and further refining is necessary to apply it to interferometric data. A year later, Serra et al. [2015] published the advanced Source Finding Application (SOFIA). The software eclipses most of its predecessors due to the broader functionality, obtained by optimizing the various methods and algorithms that were tested over the preceding years. It is the first source finding tool that incorporates a minimum of three source finding algorithms, the threshold finder; the SMOOTH PLUS CLIP source finder (S+C: Serra et al. [2012b]) and the Characterised Noise HI (CNHI: Jurek [2012]) source finder. In addition to identifying sources, it also applies parametrisation steps and returns global profiles and moment maps.

SOFIA is currently the software most large HI surveys PIs are planning to use for galaxy identify and parametrise, because of the high reliability and completeness achieved by this pipeline when tested on mock data cubes. However, it is not fully understood how SOFIA performs when applied to large interferometric datasets with real noise characteristics, this will be addressed in this thesis. I will test the performance of SOFIA on a large interferometric data cube obtained with the WSRT. Unlike with mock data cubes, the total number of genuine sources present in the cube is unknown. Hence the conventional methods of calculating both reliability and completeness are not straightforward in this case. However, there is an alternative method that we opted for in this project. We first searched the data cube using the visual method, then derived the HI parameters associated with each detection, manually. The results were then catalogued and compared with the semi-automated source catalogue by Ramatsoku et al. [2016]. The latter catalogue was produced using a combination of automated source and visual identification method (see: 4.2), hence it is referred to as semi-automated (SA) catalogue. A sample of galaxies from the visual catalogue with crossmatches from the SA catalogue was used as the testbed. We assumed that the tested sample represents the total number of genuine sources in the cube. Hence it was used to calculate both the reliability and completeness achieved by SOFIA.

This thesis is organised as follows: in Sect. 2.1 the origin of the hydrogen 21cm-line emission is discussed. Section 2.2 illustrates how this signal is captured and what type of telescope is best suited for this work. Sect. 2.3 gives an update on the current status and the future of large HI surveys. In Sect. 2.4 I review the latest top source finding algorithms that have been developed at the time of writing of this thesis. The raw HI data cube properties of WSRT HI data cube are illustrated in Ch. 3. Galaxy identification methods used in this thesis (i.e. visual, semi-automated and fully automated) and corresponding methods for calculating the HI properties of detected galaxies are described in Ch. 4. The resulting HI catalogues are presented in Ch. 5. In Ch. 5, I also discuss the measured HI parameters

and their respective characteristics. In the last chapter (Ch. 6) I present a comprehensive comparison of the three galaxy identification methods. The same chapter also includes reliability and completeness calculations. Throughout the thesis, I have assumed a Λ cold dark matter (ΛCDM) cosmological model, with matter density of $\Omega_M = 0.3$, $\Lambda_M = 0.7$ and a Hubble constant of $H_0 = 70 km\,s^{-1}\ Mpc^{-1}$.

Chapter 2

Observing the HI universe

The 1931 discovery of a radio signal at the centre of Milky Way by Jansky marked the birth of radio astronomy. The significance of this discovery was not apparent at the time and not much was known about radio wavelength. I n 1944 a Dutch astronomer predicted that HI which accounts for a large fraction of the baryonic matter can be observed in a rare transition at radio wavelengths, known as a spin-flip transition. It was not until 1951 that this transition was first observed by Ewen and Purcell [1951]. Preceding discoveries such as the cosmic microwave background (CMB) by Penzias & Wilson (1965) and the 1967 discovery of a pulsating radio star (pulsar) proclaimed radio as one of the powerful windows on the electromagnetic spectrum for probing the universe.

The radio signal has two types of emission, the continuum emission which radiates over wide range wavelengths and the spectral line emission that occurs over a specific wavelength. Continuum emission comprises of thermal emission and synchrotron radiation. Note, all hot bodies emit radio signal in a form of thermal emission. On the other hand, synchrotron radiation is found on bodies with strong magnetic fields such as the Active Galactic Nuclei (AGN), Quasars and pulsars. The spectral line is like a fingerprint for identifying atoms, elements or molecules present in the astronomical object or medium. In radio wavelength there exist over 50 spectral line[1] but the HI spectral line is the most important. The abundance of hydrogen in the universe makes HI line emission surveys a great tool for probing the universe. Secondly, the wavelength of the HI line emission is much greater than the cross-sectional area of dust grains, hence radio observations are immune to dust obscuration.

In this chapter I present the physics of the HI line emission, how can this line be captured, a process of separating it from unwanted background signals, the kind of surveys that have been done and what the future looks like for HI surveys.

2.1 The origin of the HI line emission

The hydrogen atom consists of a positively charged proton (p+) engulfed by a negatively charged electron (e^-) cloud which is held together by the Coulomb force. The hydrogen atom can only occupy discrete energy levels, and the laws of quantum mechanics also dictate that every elementary particle such as a p^+ or an e^- must carry an intrinsic spin

[1] https://www.craf.eu/iau-list-of-important-spectral-lines/

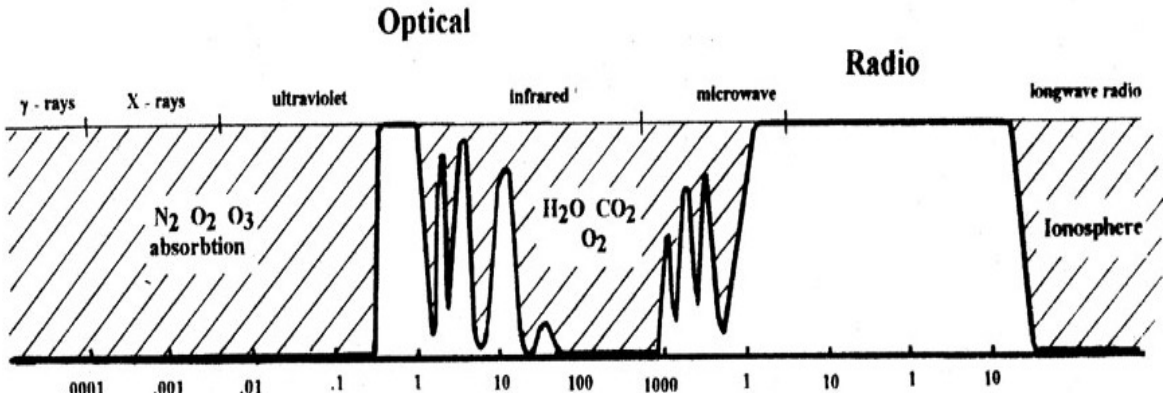


FIGURE 2.1: The electromagnetic spectrum illustrating range of wavelength windows with minimal to no atmospheric absorption. The radio window is limited by atmosphere at $\lambda \lesssim 2$ cm and ionosphere at $\lambda \gtrsim 12$ m . Credit: Burke and Graham-Smith [2009].

angular momentum. The lowest energy state for a hydrogen atom occurs when the spin of its electron is anti-aligned with the spin of its proton. However, collisions with other atoms or electrons occasionally supply the hydrogen atom with sufficient energy that might align the spin of its proton and electron. With the aligned spins, the hydrogen atom is in a slightly excited state. If left for $\sim$ 10 million years, the electron will spontaneously flip back its spin to retain the lowest anti-aligned energy state however, the abundance of HI in the Universe means this transition occurs frequently. This spin-flip transition results in emission of a photon with frequency of 1420 MHz, equivalent to a wavelength in the radio domain of $\lambda = 21$ cm.

Observations on the radio wavebands are conducted from the ground due to a wider transmission window (see Fig. 2.1). When observing galaxies in HI, the literature shows that HI disks extend much further than the optical/stellar counterpart. The consequences of a much larger HI disc are its susceptibility to tidal pulls and ram pressure stripping. However, HI observations also reveal galaxy intra-group/cluster interactions that are not notable in other wavelengths. For example Fig. 2.2 shows the M81 group in both optical (top panel) and HI (bottom panels). The optical image shows isolated galaxies with no sign of interactions, whereas HI images reveal the whole group to be one large agglomeration of gas revealing tidal interactions and HI bridges in the group.

HI line emission signals are very weak and require large telescopes with high sensitivity to be detected. In the next section, I will go over the procedure employed to detected and map HI line emission.

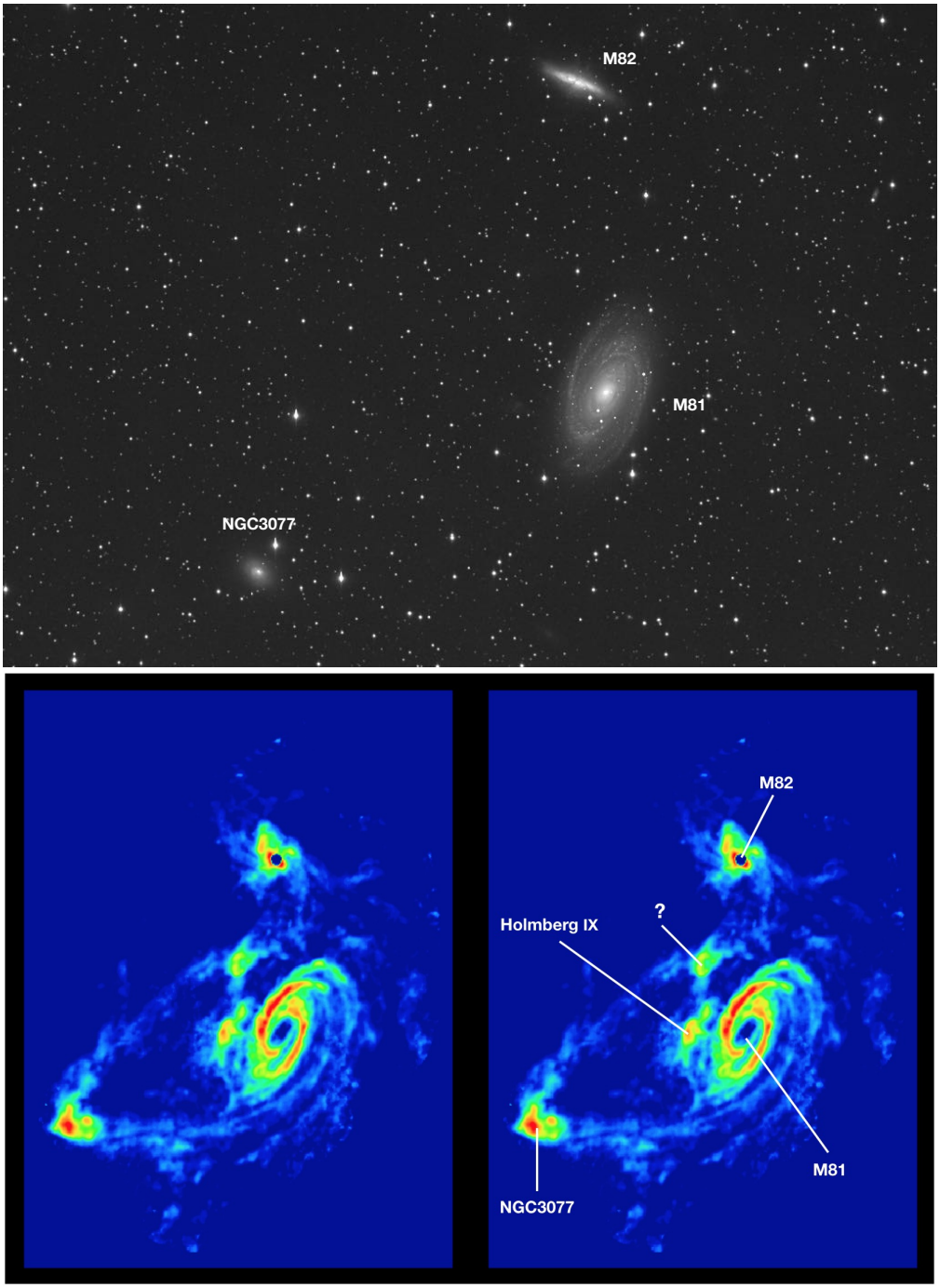


FIGURE 2.2: Optical and radio maps of the M81 group. The optical map (top panel) shows a viral stable group with no tidal interactions. Whereas the HI maps (bottom two panels) show strong tidal interaction amongst the members of the group, credit: Yun et al. [1994].

2.2 Radio interferometry

The instrument for observing HI line emission is the radio telescope with the L-band receiver. Its primary focus includes: (i) collecting large amounts of photons which can be achieved by either increasing the telescope's collecting area or integration time (or both), (ii) magnifying observed images in order to: detect fine details on resolved sources, identify distant and unresolved sources. The magnification of images is known as resolution and is mathematically defined as $\theta_R = 1.22\frac{\lambda}{D}$. Resolution in L-band is limited by the diameter/surface area of the antenna (Storey [2002]). The resolution can be maximised by increasing the diameter of the antenna. However, it is not possible to build infinitely large single dish antennas. As a result, the biggest steerable single dish antenna is the 100 m Green Bank Telescope[2] (GBT) and has a resolution of $\theta_R = 8.81$ arcminutes. There exist two more bigger single-dish radio telescopes: the 305m diameter Arecibo radio telescope[3] has a resolution of $\theta_R = 2.89$ arcminutes; and the 500m Chinese Five-hundred-meter Aperture Spherical radio Telescope[4] (FAST), with a resolution of $\theta_R = 1.76$ arcminutes. Unlike the GBT, the latter two telescopes are fixed to the ground. For greater resolution, future single-dish antennas will need much larger collecting and that does not seem feasible. Hence a different technique of achieving higher resolution and greater sensitivity such as interferometry are warranted.

Interferometry is the technique used to combine multiple single-dish antennas into a one large radio telescope with much greater sensitivity and higher resolution than single dish antenna. Unlike single dish antenna, the resolution of an interferometry is inversely proportional to the longest baseline of the array instead of aperture diameter ($\theta_B = 1.22\frac{\lambda}{B}$). This allows radio interferometry to reach resolutions that would be impossible with a single dish antenna. There are some advantages that come with using interferometry such as better pointing accuracy and susceptibility to instrument drifts in gain unlike the single dish antenna (e.g. S Stierwalt and Bemis [2016]). Since the 70's , astronomers have built a number of world class radio interferometers such as the WSRT[5], Australia Telescope Compact Array[6] (ATCA), Very Large Array[7] (VLA) and Atacama Large Millimetre Array[8] (ALMA). These telescopes have and are still contributing greatly in the quest of astronomers and cosmologists of uncovering the fundamentals of the universe (e.g. dark energy, dark matter, galaxy formation and evolution).

However, interferometry with much more superior sensitivity and resolution is required if we are to answer most of the underlying questions in both astronomy and cosmology. As a result, an enormous international project of building the world's biggest radio telescope, the Square kilometre Array[9] (SKA), is underway. Upon completion, SKA

[2] http://greenbankobservatory.org/
[3] http://outreach.naic.edu/ao/scientist-user-portal
[4] http://fast.bao.ac.cn/en/FAST.html
[5] https://www.astron.nl/radio-observatory/public/public-0
[6] https://www.narrabri.atnf.csiro.au/
[7] www.vla.nrao.edu
[8] http://www.almaobservatory.org/en/home/
[9] www.skatelescope.org

will have a square kilometre of collecting area; a maximum baseline of about 3 000 km [Taylor, 2012]. Due to the scale of the SKA project, a number of precursors and pathfinders are being constructed around the globe, which will be used as testbeds for both scientific and engineering purposes. In the next section, I will discuss some of these precursors and highlight large HI surveys planned for these pathfinders.

FIGURE 2.3: The Karoo Array Telescope (MeerKAT) with the Gregorian offset. Credit: SKA-SA

2.3 The SKA precursors

In preparation for the SKA, precursor telescopes such as the South African Karoo Array Telescope[10] (MeerKAT, see Fig. 2.3), the Australian SKA Pathfinder[11] (ASKAP), the Murchison Widefield Array[12] (MWA) and the Hydrogen Epoch of Reionization Array[13] (HERA). In addition to the four precursors, there exist a number of international pathfinder telescopes and instruments dedicated to SKA preparations (e.g. APERture Tile In Focus (the APERTIF, Verheijen et al. [2008]), Arecibo observatory, Parkes telescope, etc[14]). The precursors and pathfinders are and will conduct surveys related to the SKA future projects in order to assist in developing and testing equipment and pipelines required for the SKA. In this section I will highlight large HI surveys (LHIS) planned or currently under way on two precursors and one pathfinder.

Planned large HI surveys

About 70% of MeerKAT observing time is dedicated to ten Large Survey Science Projects (LSSPs)[15] which have PIs from across the globe including South Africa. Three of these LSSPs are large HI surveys: the Looking at the Distant Universe with the MeerKAT Array (LADUMA) which will probe the HI line emission in the single light cone extending to $z = 1.4$ [Baker et al., 2018], the MeerKAT HI Observations of Nearby Galactic Objects

[10] http://www.ska.ac.za/gallery/meerkat/
[11] http://reionization.org/
[12] http://www.mwatelescope.org/
[13] http://reionization.org/
[14] https://www.skatelescope.org/precursors-pathfinders-design-studies/
[15] http://public.ska.ac.za/meerkat/meerkat-large-survey-projects

Observing Southern Emitters (MHONGOOSE) which is a deep survey of HI distribution in a sample of 30 nearby disk and dwarf galaxies [de Blok et al., 2017], and the MeerKAT HI Survey of the Fornax Cluster covering a mosaic area of 12 $\deg^2$, planned to study galaxy-galaxy interaction and galaxy intra-cluster medium interaction [Serra et al., 2017].

Upon ASKAP completion, 75% of the observing time in the first 5 years will be dedicated to conducting 10 approved LSSPs[16]. Amongst these LSSPs, there are two LHIS: 1) the Widefield ASKAP L-Band Legacy All-Sky Blind Survey (WALLABY), which is an extragalactic HI survey covering 75% of the sky ($-90° < \delta < +30°$) over a redshift range of 2 000 to 77 000 $\mathrm{km\,s^{-1}}$ and is estimated to identify over 500 000 galaxies [Duffy et al., 2012], 2) the Deep Investigations of Neutral Gas Origins[17] (DINGO), which will study HI evolution out to a redshift of $z < 0.4$ [Meyer, 2009].

It is apparent that HI surveys will play a major role in attempting to answer some of the fundamental questions in the field of astronomy and physics. However, HI surveys were not always considered as viable tools for studying and understanding the big questions in astronomy. In the next chapter, I will give an overview of the rise of HI surveys in the field of astronomy.

The rise of HI surveys

Before the 90's, extragalactic blind HI surveys were not as common as the present era. HI surveys were traditionally conducted as follow-ups to optical target samples. These surveys were generally motivated by scientific goals other than understanding the characteristics of an unbiased HI source population. HI surveys were used as tools to research questions such as environmental impact on gas rich objects found in clusters, constraining cosmological distance scale parameters Giovanelli and Haynes [2016]. Two surveys, the Arecibo HI strip Survey (AHISS, Zwaan et al. [1997]) and the Arecibo Dual-Beam Survey (ADBS, Rosenberg and Schneider [2000]) were the first to provide insight on the statistical significance of dwarf galaxies. The HI Parkes All Sky Survey (HIPASS, Meyer et al. [2004]) was the first wide-field extragalactic HI survey. It covered $\sim$ 75% of the whole sky (i.e. 30 000 $\deg^2$) and it detected about 5 000 galaxies. In 2005 the Arecibo L-band Feed Array (ALFA) become available at the 305 m Arecibo telescope and it led to a number of extragalactic HI surveys. The surveys included the Arecibo Legacy Fast ALFA survey (ALFALFA, Giovanelli et al. [2005]) covering an area $\sim$ 6 920 $\deg^2$ with 31 000 source detections and ALFA Zone of Avoidance Deep Survey (ALFA-ZOA, Henning et al. [2008]), covering $\sim$ 300 $\deg^2$ and detecting 1 200 sources (see McIntyre et al. [2015]). Both the ALFALFA and the ALFA-ZOA surveys cover a significantly much smaller area compared to HIPASS, but they are deeper in redshift.

[16] https://www.atnf.csiro.au/projects/askap/ssps.html
[17] http://internal.physics.uwa.edu.au/ mmeyer/dingo/project.html

The ALFA-ZOA survey was aimed at mapping out the connections between large scale structures across the Galactic plane. Systematic blind HI surveys have been used extensively in the ZoA since it is the only method that allows mapping of the large-scale structures (LSS) through the Milky Way as it's not susceptible to dust obscuration or light pollution. The Dwingeloo Obscured Galaxy Survey (DOGS) is one of the first early HI blind surveys of the ZoA, it covers the northern ZoA ($|b| < 5$ deg) but had a low sensitivity and was also shallow in redshift ($0 < cz < 4\,000$ km s^{-1}) Kraan-Korteweg et al. [1994]. The Parkes multi-beam receiver has played a significant role revealing the extragalactic structures hidden by the southern Galactic plane. The early surveys conducted by this instrument include the southern hemisphere Parkes HI ZoA Survey (HIZOA-S, Staveley-Smith et al. [2016]) and its northern extension (HIZOA-N, Donley et al. [2005]). HIZOA-S covers a Galactic longitude range of $212^\circ < \ell < 36^\circ$, Galactic latitude $|b| \leq 5^\circ$ and has identified 833 galaxies to a recessional velocity of 12 000 km s^{-1}. Whereas HIZOA-N surveys has a Galactic longitude range of $36^\circ < \ell < 52^\circ$ & $196^\circ < \ell < 212^\circ$, Galactic latitude $|b| \leq 5^\circ$ and detected 77 galaxies of which 20 had been previously detected. These surveys including the ALFA-ZOA have contributed in exposing the gas-rich extragalactic structures in the ZoA but left a massive part of ZoA in the northern sky ($80^\circ < \ell < 180^\circ$) unmapped. So far, these large area surveys have been performed by single-dish antennas.

However there are also large HI surveys that have been conducted using interferometry, such as the Westerbork HI survey of irregular and spiral galaxies (WHISP, van der Hulst et al. [2001]), blind survey of nearby spiral-rich Ursa Major Cluster (VLA, Verheijen et al. [2000]), and the blind survey of Canes Venatici region using WSRT (CVn, Kovač et al. [2009]), the HI blind survey of Perseus-Pisces in the ZoA using the WSRT (HIPP-ZoA, Ramatsoku et al. [2016]). HIPP-ZoA survey is a follow-up survey in the vicinity of a hypothesised Perseus-Pisces ZoA crossing (see Focardi et al. [1984]). This study investigated the structure associated with the 3C 129 galaxy cluster and the environmental effects on HI properties of the identified galaxies. The resulting HI data cube is ideal for testing and characterising data handling algorithms/pipelines such as the fully automated source finding and parametrisation pipelines.

2.4 Detection and parametrisation of HI line emission

In radio astronomy, a source finder is a computer program that uses an algorithm to search radio data cube for astronomical objects, then returns a source catalogue comprising of sky coordinates and redshifts of the detections. In order to derive other source parameters (e.g. integrated flux and HI mass), further parametrisation steps are necessary.

2.4.1 Source finding algorithms

I have chosen five source finding algorithms that are likely to be incorporated to various source finding and parametrisation tools, for use in the large HI surveys: 1) GAMMA-FINDER [Carilli and Rawlings, 2003], 2) the SMOOTH PLUS CLIP (S+C) finder [Serra et al.,

2012a] 3) DUCHAMP [Whiting, 2012], 4) the Characterised Noise HI (CNHI) source finder [Jurek, 2012], 5) the 2D-1D WAVELET RECONSTRUCTION source finder [Flöer and Winkel, 2012]. Here I provide a brief description for each of the above source finders. For a more in-depth review of these source finders, I refer to the reference papers.

1. **GAMMA-FINDER**: is based on the Gamma-Test [Stefánsson et al., 1997] which estimates a Gamma-statistic (Γ) by measuring noise variations across the data cube. Γ values reveal segments of data distinct from the global noise distribution. Such segments corresponds to potential sources. A user can specify signal-to-noise ratio and use it to qualify or reject segments as genuine detections. GAMMA-FINDER only returns source spatial coordinates and redshift but does not parametrises sources.

2. **S+C finder**: makes use of a limited number of filters to optimise the signal-to-noise ratio of all sources present in the data. S+C finder searches for line emission in multiple spatial and spectral resolution by applying three-dimensional smoothing kernels (specify by the user) to the original data cube. At each specific resolution, voxels with an absolute flux value above a user-specified flux threshold are detected and saved into sub-mask cubes. The union of all constructed sub-mask at different resolution make up the cube mask. The cube mask contains all the sources detected by S+C finder.

3. **DUCHAMP**: is a source finder intended for three-dimensional data, however, it can be applied to two-dimensional or one-dimensional data. This finder identifies sources using two steps: 1) applies a flux or signal-to-noise threshold then identifies all sources above the threshold, 2) rejects spurious based on criteria specified by the user and merges genuine emission detections into sources. DUCHAMP offers the following functionalities which can improve pipeline performance: data pre-conditioning; spectral & spatial smoothing of data cube before search; application of filters and wavelet reconstruction of the entire data cube. The last step is parametrisation, where DUCHAMP measures HI parameters associated with each source detection.

4. **CNHI source finder**: sees data cube as a collection of spectra. It employs the Kuiper's test (1960) to identify spectrum segments that are not consistent with the pure noise spectrum. This is achieved by calculating the probability that the test segments and the rest of the spectrum originate from the same class of voxels. if the probability is low, then the segments are classified as spectra of genuine sources in the data cube. Note, the probability cut/threshold is defined by the user. Once all the spectra have been dealt with, the source segments are combined into sources using an extended version of Lutz's one-pass algorithm [Lutz, 1980]. This source finder assumes that the spectra are dominated by noise, which is a valid assumption for spectral data cubes. However, for Kuiper's test to work effectively, the test segment must be at least four channels wide. This may lead to misidentification of genuine sources with low linewidths as false positives.

5. **2D-1D WAVELET RECONSTRUCTION source finder**: is an adaptation of multi-dimensional wavelet denoising scheme [Starck et al., 2009]. The finder performs a two-dimensional wavelet transform in all planes of the data cube, then performs a one-dimension wavelet transform along every line of sight. A threshold is set by the user to determine significant wavelet coefficients which construct a noise-free data cube. The resulting data cube can be used as a mask for source detections in the original data cube.

Popping et al. [2012] conducted a performance study by comparing the five aforementioned source finders. They applied each source finder to 2 three-dimensional spectral cube models. The first cube contained point sources with narrow Gaussian profiles, while the other had spatially extended galaxies with inclination and rotation curves. They found that all the source finder performed relatively well, however, DUCHAMP gave the best results in terms of completeness when it comes to point sources, whereas S+C finder showed best results for extended sources. Other findings showed that S+C finder best represent the three-dimensional structures of the detected sources.

Partly based on these findings, the S+C finder is one of the leading source finder considered for SKA pathfinders' large HI survey. However a source finders account for identifying sources, once identified, parametrisation steps are to follow. Hence the need for advance automated parametrisation pipeline. A pipeline that offers both source finding, and parametrisation algorithms would be ideal for the upcoming large HI surveys. Such pipeline exists and is known as SOFIA. An overview of SOFIA is presented in the next section.

2.4.2 Overview of SOFIA

SOFIA is a flexible pipeline for identifying and parametrising sources in a three-dimensional spectral data cube. It comprises of one basic source finder (i.e. FLUX THRESHOLD FINDER), two advance source finding algorithms (i.e. CNHI & SOFIA) and series of parametrisation algorithms. Here I will present an overview of SOFIA as it one of the three source finding methods used in this project.

A flow chart highlighting various modules available in SOFIA is presented in Fig. 2.4. The flowchart is arranged in a clockwise manner, with DATA PRECONDITIONING being the first step, while PIPELINE OUTPUT is the last. Once a data cube or sub-volume is loaded into SOFIA, the modules allow the user to prepare the data for SOFIA smooth run, detect spectral line emission, merging detected voxels into sources, reject spurious sources, apply mask optimisation to individual detections, measure source parameters and derive source outputs (i.e. moment maps & integrated spectra).

1. DATA PRECONDITIONING

   - Input data cubes − comprises of spectral line emission convolved with instrumental noise and errors.

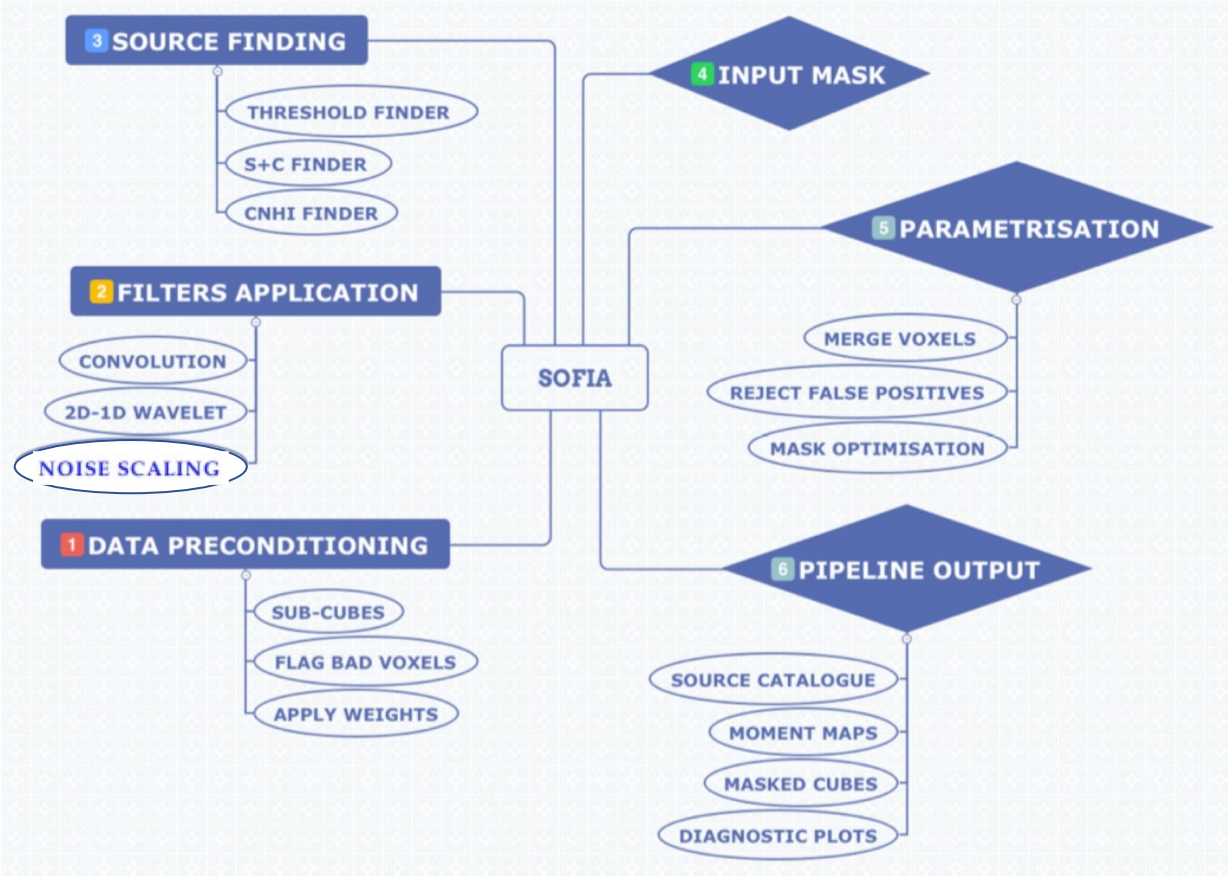


FIGURE 2.4: The flow chat of SOFIA, highlighting all the steps to be taken for successful run of SOFIA.

- Weight functions/cubes – allows a user apply weights to take into account noise variations across the cube and flag regions with imaging artefacts.
- Flag bad voxels – flag regions with RFI contamination.

2. FILTER APPLICATION

- SOFIA offers two filters: i) convolution with a three-dimensional kernel (defined by the user), ii) the 2D-1D wavelet de-noising algorithm by Flöer and Winkel [2012]. These algorithms return a noise-free cube reconstructed from wavelet coefficients above a user-specified threshold.
- An optional control parameter that normalises noise levels across the spectral cube. There are two ways of normalising noise, i) local: measurements of the noise occur in a running window specified by the user and on a specific grid,

ii) global: measurements are conducted on an image plane perpendicular to the axis on which noise is to be scaled.

3. SOURCE FINDING
   SoFiA offers three unique source finders: THRESHOLD, CNHI and S+C finder. The THRESHOLD finder prompt the user to set an n-sigma ($n \in \mathbb{R}$) flux detection threshold and SoFiA will return all voxels above n-sigma. A combination of source finders and filters can in theory form alternative source finders (e.g. combination of the 2D-1D wavelet filter & THRESHOLD finder).

4. INPUT MASK
   An optional file which contains a mask of pixels identified as part of a source (this could be from a previous run of SoFiA or other software). This functionality allows for a re-parametrising sources without repeating the source finding step.

5. MERGING DETECTED VOXELS
   All SoFiA source finders return binary mask. The binary mask contains the basis for establishing individual sources. This establishment is carried out using the C++ implementation of Lutz [1980] one-pass algorithm assembled by Jurek [2012]. This algorithm mimics that of the well-established friend-of-friends algorithm. At this stage, all the individual sources have derived basic parameters (i.e. normalised total flux and peak flux).

6. RELIABILITY CALCULATION
   Note that all the source finders require the user to specify a detection threshold. The closer this threshold is to the cube's noise level, the more noise peak will be identified and included in the final binary mask Serra et al. [2015]. SoFiA possess two methods of calculating reliability. The first method is size filter, based on the factor that all true detections are at least as large as the cube's resolution. The downside of this method is that it may flag out bright unresolved genuine sources that do not meet the size threshold given by the user. The second method quantifies reliability of each detection by comparing the distribution positive flux sources with negative flux sources. This method is based on two assumptions, firstly, the noise within the data cube is symmetric (i.e. the total flux distribution of detection with positive is similar to that of detections with negative). Secondly, all positive (genuine) detections have positive flux only and do not have negative flux counterparts. This method is explained in great depth by Serra et al. [2012a].

7. MASK OPTIMISATION
   SoFiA calculates HI parameters (e.g. total flux, linewidths, etc.) using detected voxels in the final binary mask. However, literature cautions that masks can miss relatively faint emission and outer edges of sources. This will consequently result in underestimating both total flux and size of the source, see Westmeier et al. [2012]. To correct for this phenomenon, SoFiA offers two mask optimisation methods that grow the masks of individual detections (independently of other detections). The

first method fits an ellipse on a moment zero map. This ellipse is then used as an initial mask for all spectral channels spanning the source, hence resulting in a cylindrical mask. The cylindrical mask is then grown until maximum total flux of the source is achieved. This method is well suited for resolved and face-on detections. The second method involves growing the initial mask on the two spatial axes of the data cube. The size of the structural element is iteratively increased until total flux of the source converges. This method preserves the 3D shape of the source and also achieves more accurate flux measurements.

8. SOURCE PARAMETRISATION
   After the mask optimisation step, SoFiA recalculates all the HI parameters such as flux centroid; geometrical centroid; minimum and maximum voxel value; linewidths derived using different conventional methods; source size and boundaries; results of a busy function to integrated spectrum, see Westmeier et al. [2012] for details; results ellipse fit to the moment zero map. These parameters are measured in both raw and WCS coordinates.

9. PIPELINE OUTPUT
   After a successful run, SoFiA returns a source catalogue with 58 source parameters. However, it allows a user to calculate only the parameters of interest. It also returns three data products: mask, moment zero map and a moment one map.

In this chapter, I have discussed how the neutral hydrogen line emission is detected, mapped and parametrised. I also gave an overview of the source finding and parametrisation pipeline developed for working with large interferometric spectral cubes (i.e. SoFiA). The HI data cube from the HIPP-ZoA is ideal for conducting a performance study of SoFiA and contrast it with the traditional methods of source identification and parametrisation (i.e. Visual and Semi-automated methods).

Chapter 3

Data acquisition

Here I present the observing set-up of the HIPP-ZoA survey. The data cube used in this project resulted from this large interferometric survey.

3.1 HIPP-ZoA survey parameters

The HIPP-ZoA survey was carried in between August to November 2012 with the WSRT in the Netherlands. The WSRT consists of 14 antennas with a diameter of 25 meters per antenna. The survey was carried out in a hexagonal mosaic configuration of 35 pointings, with a separation of 0.5 deg. Each pointing had an integration time of 12 hours (2×6 hrs). The survey covers a field of view of 9.6 sq.deg, spans a redshift range of $2\,400 - 16\,600$ km s^{-1}, has a channel width of 8.24 km s^{-1} , a root mean square (rms) noise level of $\sigma_{\rm rms} = 0.4$ mJy beam^{-1} and an angular resolution of 23×16 arcseconds. The surveyed area contains the rich X-ray galaxy cluster 3C 129 (see Krawczynski [2002]). The survey is wide enough for mapping possible interactions between the galaxy cluster 3C 129 and the Perseus-Pisces filament.

In Table 3.1, I present a summary of these parameters. Individual pointing parameters can be found in Table 1 of Ramatsoku et al. [2016].

TABLE 3.1: The observing parameters of the HIPP-ZoA survey are reproduced.

Parameter	Value	Units
Sky coverage	9.20	Sq.deg
Redshift (cz) range	$2\,400 - 16\,600$	km s^{-1}
Survey centroid	($\ell = 160.52$, $b = 0.28$)	deg
Beam size ($\theta_{\rm R_{res}}$)	23×16	arcsec^2
Channel width	8.24	km s^{-1}
Root Mean Square Noise ($\sigma_{\rm rms}$)	0.40	Jy beam^{-1}
Integration time	2×6	hrs. pointing^{-1}

Note, data reduction procedure (i.e. RFI flagging; phase & amplitude calibration; data cube construction; continuum subtraction; image/cube cleaning and Mosaicking) is presented in Sect. 2.1 of Ramatsoku et al. [2016] and since there are no adjustments made, its inclusion to this thesis is not necessary. Upon completion of the data reduction steps, a HIPP-ZoA data cube was assembled. HIPP-ZoA data cube is made up of nine subcubes

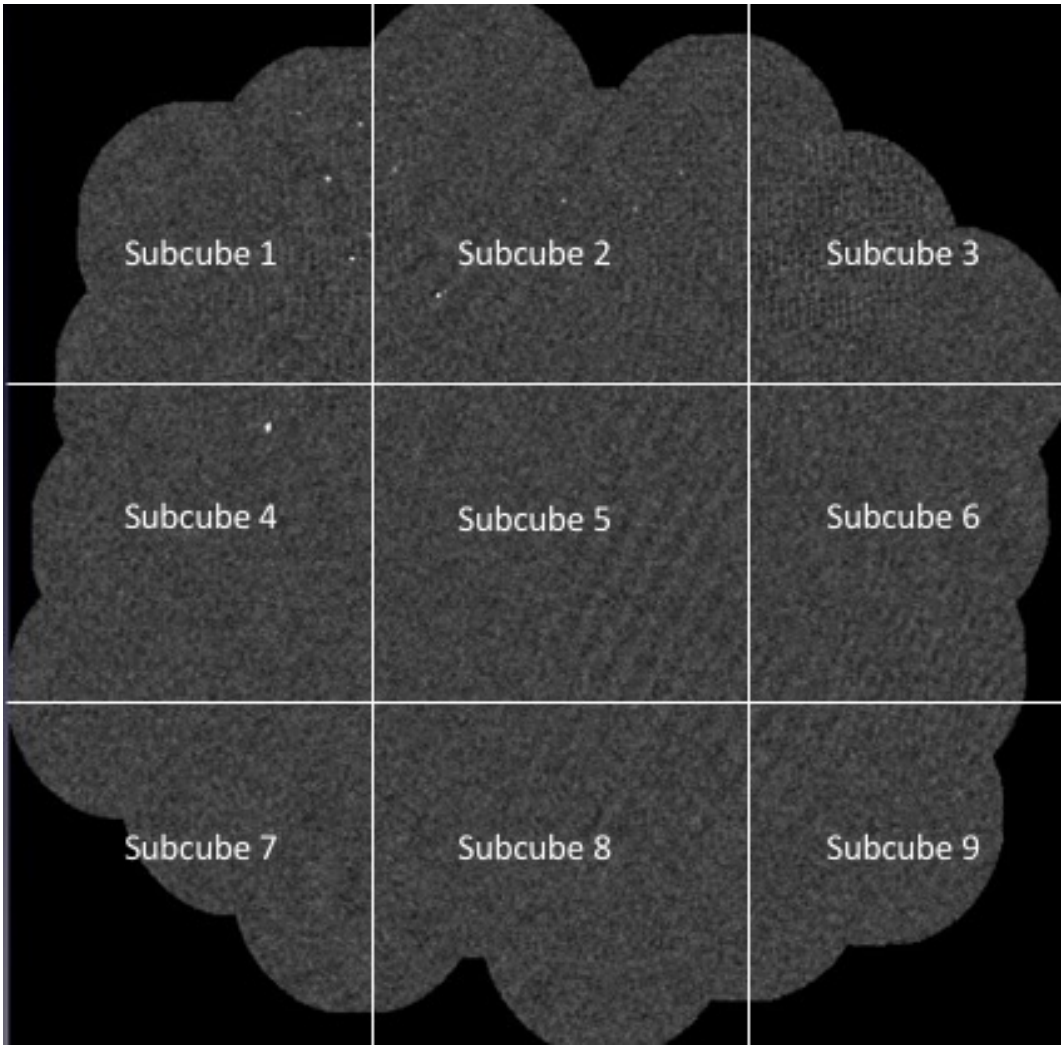


FIGURE 3.1: In grey is the projected sky distribution of the observed mosaic (from Ramatsoku et al. [2016]). The while solid lines mark subcubes of HIPP-ZoA data cube.

(see Fig. 3.1). Each subcube spans the entire redshift range of the survey (i.e. 1717 channels) but has spatial dimensions of $\sim 750 \times 750$ sq.pixels. Contiguous subcubes have spatial overlap of ~ 0.06 deg (33 pixels). Sources on the overlap regions will be contained on both contiguous subcubes. This will be useful when doing source finding and attempting to correct for source truncation due to subcube boarders.

Chapter 4

Methodology

Chapter 3 presented the observational procedure employed on the HIPP-ZoA survey to produce the HIPP-ZoA data cube. In this chapter, I describe how this data cube was inspected for galaxy candidates using three unique methods: Visual inspection, semi-automated and fully automated method. I detail how each inspection/ source finding method was carried out and the procedure followed in extracting HI source parameters such as flux centroid, linewidths, integrated flux, total HI and etc.

4.1 Visual-inspection

Here I describe the manner in which a visually inspected source catalogue was derived. It is a process of three different steps. Firstly, I used a visualisation tool to identify and record each and every galaxy candidate's flux weighted centroids. I then extracted a subcube for each galaxy candidate and derive other HI parameters. Lastly, I produced moment maps associated with each candidate and measured flux weighted centroid.

4.1.1 Source identification

The data visualisation is carried out using KVIS, a three-dimensional visualisation tool from KARMA Gooch [2006]. KVIS allows the user to view the data cube in three different planes while stepping through a third axis. This is ideal for this work because galaxy morphology is strongly dependent on the observing plane. For example, if an unresolved galaxy is viewed on the spatial plane (i.e. XY-plane), it appears as a point source, whereas on a spectral-spatial plane (i.e. XZ/YZ-plane) it will be slightly elongated. Bright galaxies with significantly small linewidths can appear as noise peaks when viewed on a spectral-spatial plane (i.e. XY-plane). Figure 4.1 shows channel maps of two galaxies viewed in three different planes. The ellipses/circles enclose HI emission of individual galaxy candidates. The top panels show two galaxies (i.e. galaxy A and galaxy B) as seen in XY-plane at two frequencies: $\nu_1 = 1\,396.967$ MHz and $\nu_2 = 1\,396.225$ MHz. The middle panels show the two galaxies in the XZ-plane at their respective declinations: $\delta_1 = 44^\circ\,35'\,36.29''$ and $\delta_2 = 44^\circ\,36'\,30.39''$. The bottom panels present the two galaxies as seen in the YZ-plane at right ascension values: $\alpha_1 = 04^{\rm h}\,41^{\rm m}\,51.88^{\rm s}$ and $\alpha_2 = 04^{\rm h}\,51^{\rm m}\,3.29^{\rm s}$. Unresolved galaxies appear as dots on the XY$-$plane and similar to noise peak, however, they appear slightly

elongated on the YZ-plane and XZ-plane), whereas noise peaks still appear as dots. It is that reason I decided to view the data cube in all planes.

In order to be sensitive to all types of galaxies in the HIPP-ZoA data cube (i.e. from dwarf irregulars to massive spirals), I have opted to search for galaxies in all three observing planes. The following schematic was followed when identifying the galaxy candidates:

1. Load the data cube into KVIS
2. Choose the observing plane (e.g. XY-plane)
3. Find a relatively faint galaxy from the cube, then fine-tune KVIS settings to facilitate the identification of fainter sources.
4. Run KVIS over 100 channels/pixels at a time then move to the next 100 channels/pixels. Ensure a 20 channel/pixel overlap between the runs so not to miss smaller galaxies.
5. Loop KVIS over the first bin (B_0) and identify galaxy detections.
6. Record each detection's central coordinates (i.e. RA, Dec & Velocity).
7. After satisfactory search of B_0 move to its neighbouring bin (B_{n+1}) and repeat step 5 and 6.
8. Apply step 7 for all n, where n is the total number of bins.

This resulted in three independent galaxy candidate lists (i.e. XY, ZY and XZ), with each list corresponding to one viewing plane. The three lists were then reconciled and merged into one list. There are no prior source identification criteria that I followed but sources were accepted as candidates if they were visible in more than one plane and were slightly extended in velocity. Uncertain detections were revisited and viewed in all observing planes. Moment maps and global profiles were constructed for all the galaxy candidates. In Sect. 4.1.2 and 4.1.3, I elaborate on how these steps were carried out.

4.1.2 Data products of identified galaxies

Here the procedure for deriving data products such as moment maps, position-velocity slices and global profiles is presented. A well thought systematic process is essential when producing these maps as they have direct implications for the precision of measure HI parameters. This is because the derived HI parameters such as S_{int} and HI are based on these maps.

HI galaxy sub-volumes

The first task after completion of the detections' or galaxy candidates' list, was to extract sub- volumes enclosing each detection where possible. Using sub-volumes instead of the entire cube minimises a number of challenges such as source confusion and contamination of local background noise due to emission from neighbouring sources. However,

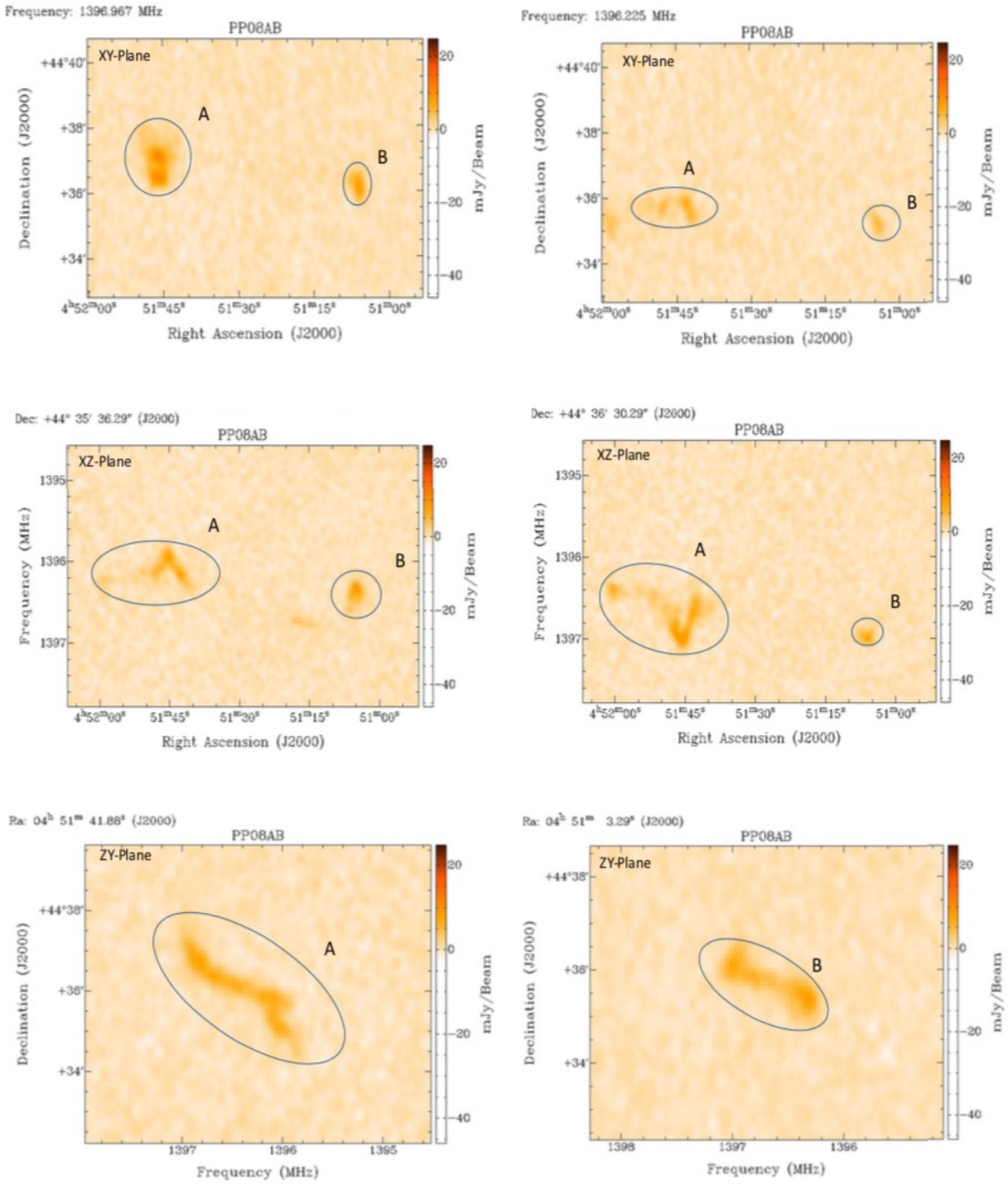


FIGURE 4.1: Two galaxies (A and B) as seen in three viewing planes. The top panels: channel maps as seen in the XY-plane. Middle panels: position-velocity slices viewed from the XZ-plane. Bottom panels: position-velocity slices viewed from the ZY-plane

this is a delicate step and it has to be handled with care because an error in the coordinate transformation from the parent cube to the sub-volume could lead to inaccurate spatial coordinates and redshift. Keeping that in mind, I wrote a script that extracted sub-volumes cantered on each detection with spatial dimensions of 100×100 sq.pixel and 200 channels spectrally. These dimensions ensure that the entire HI emission of the source detection is

accounted for. Post visual inspection were performed to ensure the success of the coordinate transformation.

The HI global profiles

The HI global profiles were created from the raw sub-volumes by using the MBSPECT module from the MIRIAD package by Sault and Killeen [1996]. MBSPECT module has a range of polynomials from first-order to higher for baseline fitting, which a user can optimise depending on the noise level in the sub-volume. To get the best fitting baseline, the sub-volumes were set to span at least 500km s^{-1} from each end of the double or Gaussian shaped profile. The profile and any nearby sources which appear along the spectral axis are masked out so not to bias the noise baseline fit. In addition, a Hanning-smoothing of the order 3 is applied.

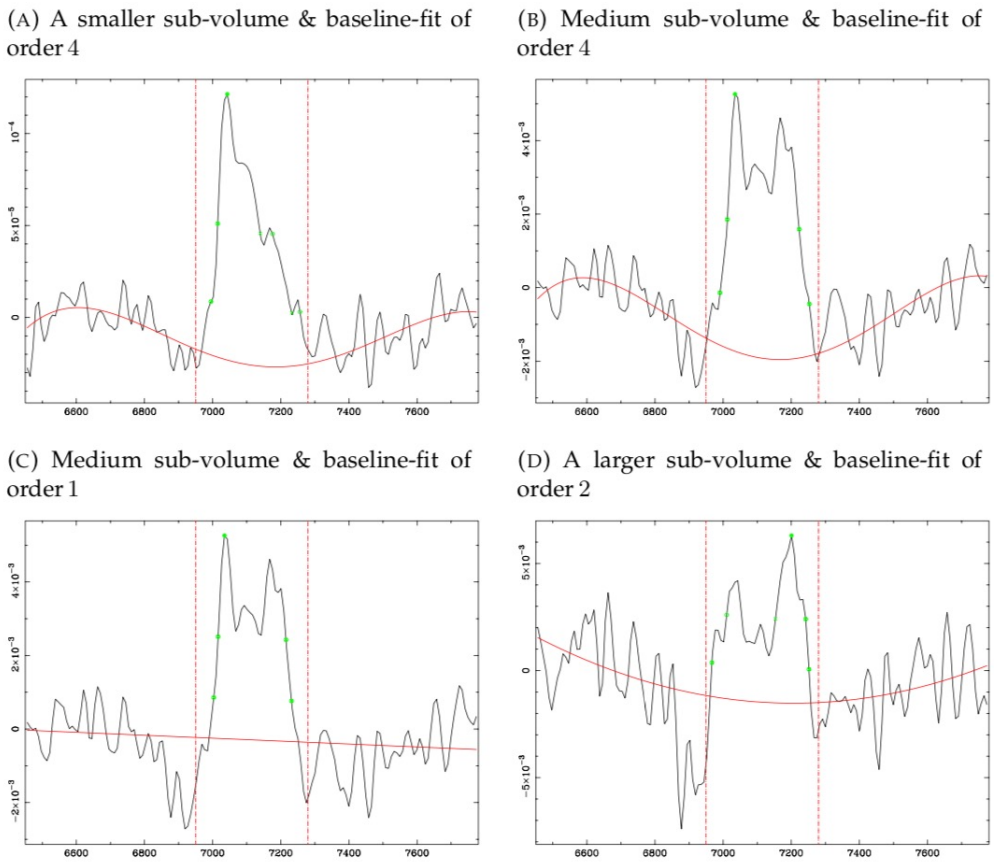


FIGURE 4.2: HI global profiles of one galaxy derived from sub-cubes of different sizes. The parallel lines mark the profile. The horizontal solid line marks the noise baseline.

Shown in Fig. 4.2 is the graphical presentation of steps taken to arrive at final acceptable HI global profile. Sub-figure 4.2a fits a perfect noise baseline of order 4 but is using an insufficient extent of the sub-volume, hence the profile fails to account for all the HI line emission and only one horn is revealed. Sub-figure 4.2b is based on a sub-volume that captures all the galaxy emission. Sub-figure 4.2c fits a first-order polynomial baseline which is a good approximation but will result to slight under-estimation of HI parameters.

Sub-figure 4.2d shows an over-fitting scenario, where a large extent of the sub-volume is applied, and baseline fit does not best represent the data. The latter scenario has major implications as it shows the profile to be buried in the noise. So the best MBSPECT parameter combination for this galaxy are those presented in sub-fig 4.2b. This is the procedure that was used to derive all the profiles presented in this thesis. It must be noted that for every source detection, different order of polynomials were tried in order to achieve possible global profile for each detection. Also the size of the sub-volume enclosing the source detections is dependent on the HI extent of each detection.

The HI moment maps

The moment zero map (M_0) of a galaxy is given by the product of the channel width (dv) and a sum of all the spectral line emission along the spectral axis,

$$M_0 = dv \sum_{i=0}^{N} S(v_i) \tag{4.1}$$

where: dv is the channel width in km s^{-1}; i marks the i^{th} voxel ($i \in [1, 2, 3, ..., N]$); N is the total number of spectral channel and $S(v_i)$ is the flux density in Janskys.

This is equivalent to collapsing the spectral sub-volume along the spectral axis resulting into a two-dimensional image of the target galaxy. It must be noted that masks are applied to all the sub-volumes before deriving the M_0 maps. These masks were created using specialised python scripts and are dependent on HI extent. An example of a moment map of one of the massive galaxies from the HIPP-ZoA data cube is presented in Fig. 4.3. Majority of the identified galaxies in this thesis are unresolved, hence their respective moment maps are not as detailed as this one.

4.1.3 HI parametrisation

I have presented the criteria and procedure for galaxy identification and production of corresponding maps. The next crucial step is the HI parametrisation. Where the global profiles are utilised to derive HI parameters such as the linewidths at 20% and 50% of the peak flux (W_{20} and W_{50}, respectively), the peak flux ($S_{\rm peak}$), the heliocentric velocity ($V_{\rm hel}$) and the integrated flux ($S_{\rm int}$).

Derived HI parameters

Figure 4.4 presents a sketch of a global profile (solid black curve). The baseline is marked by the line. Marked on the profile is the peak flux ($S_{\rm peak}$) and the velocities ($V_{\rm r,20}$, $V_{\rm a,20}$, $V_{\rm r,50}$, $V_{\rm a,50}$) corresponding to the 20% and 50% of the peak flux. It should be noted that all the HI parameter measurements are performed from the baseline subtracted profile.

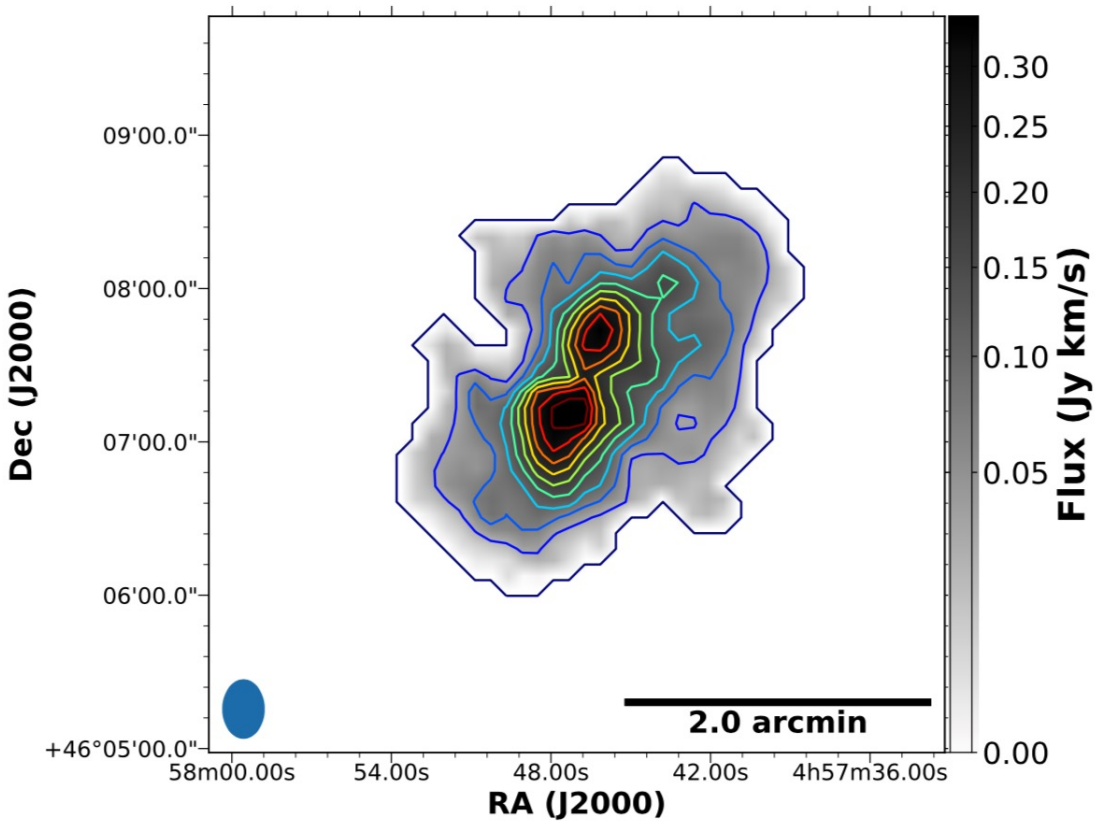


FIGURE 4.3: A moment zero map of one of the massive galaxies (J045811.74 + 460226.5) from the HIPP-ZoA data cube. The grey diffused grey emission marks integrated flux. The contours show flux gradient and, on the bottom left corner is the primary beam.

The heliocentric velocity is determined according to:

$$V_{\rm hel} = \left(\frac{V_{\rm r,50} + V_{\rm a,50}}{2} \right) \tag{4.2}$$

The linewidths of the profile are given by:

$$W_{20} = V_{\rm r,20} + V_{\rm a,20} \tag{4.3}$$

$$W_{50} = V_{\rm r,50} + V_{\rm a,50} \tag{4.4}$$

The total integrated flux ($S_{\rm int}$) in Jy m s^{-1} is derived by integrating the profile according to:

$$S_{\rm int} = \int S_{\rm v} dv \,, \tag{4.5}$$

where $S_{\rm v}$ is the measure of flux density [Jy] at each channel and dv is the channel width in units of km s^{-1}.

These are the parameters measured directly from the global profiles. However an additional three important parameters namely the signal-to-noise ratio (SNR), the cosmological distances to the galaxies and HI masses can be derived from the combination of

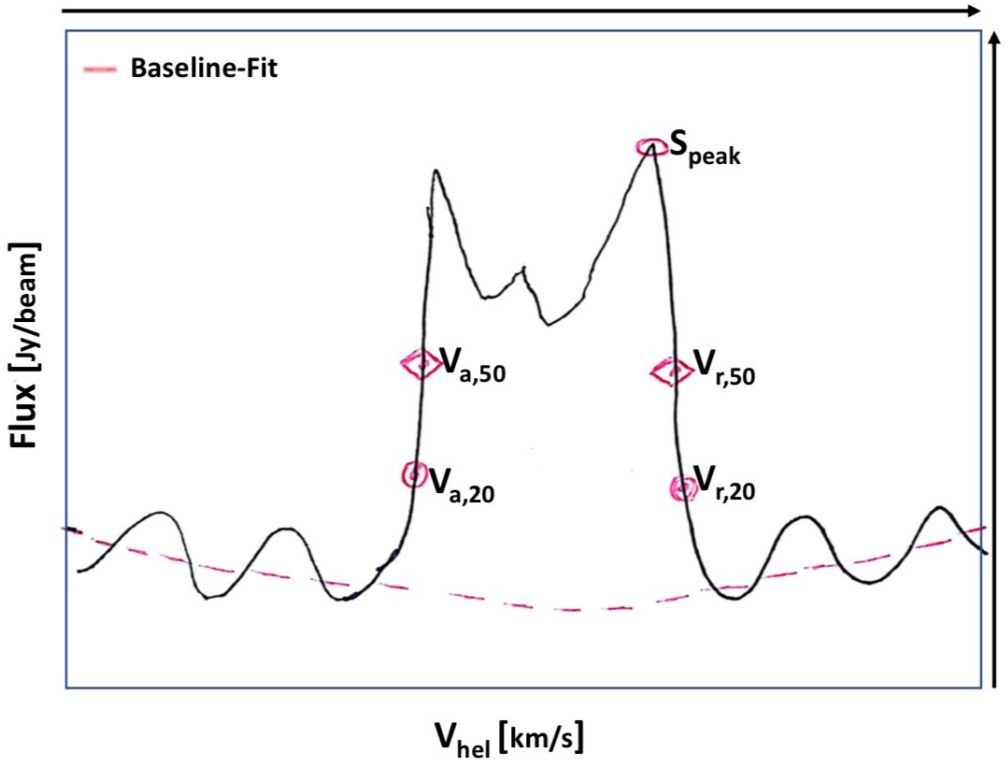


FIGURE 4.4: The sketch of an HI global profile illustrating measure HI profiles.

measured parameters.

Following Saintonge [2007] and Cortese et al. [2008], the signal-to-noise ratio is derived by:

$$SNR = \left(\frac{S_{\rm int}\sqrt{w}}{\sigma_{\rm rms} \times W_{50}} \right) , \tag{4.6}$$

where w is given by $W_{50}/(2dv)$ for W_{50} less than $400\,{\rm km\,s^{-1}}$, or $(400\,{\rm km\,s^{-1}})/(2dv)$ for W_{50} of $400\,{\rm km\,s^{-1}}$ or greater. It must be noted that $S_{\rm int}$ is measured in [Jy km s^{-1}] and the spectral cubes' $\sigma_{\rm rms}$ is in [Jy] in this equation.

The cosmological distances, D in Mpc, are given by the Hubble equation:

$$D = \frac{V_{\rm hel}}{\rm H_0}, \tag{4.7}$$

where $\rm H_0$ is the Hubble constant with the units of km s^{-1} Mpc^{-1}.

It should be noted that no corrections were applied for bulk flows and other extragalactic or cosmological phenomenon that can bias these the cosmological distances.

The final parameter is the total HI mass ($M_{\rm HI}$) in units of solar masses ($M_\odot$) and is calculated according to:

$$M_{\rm HI} = \left(2.36 \times 10^5\right) D^2 S_{\rm int}\,, \tag{4.8}$$

where D is measured in Mpc, $S_{\rm int}$ is in Jy km s^{-1}.

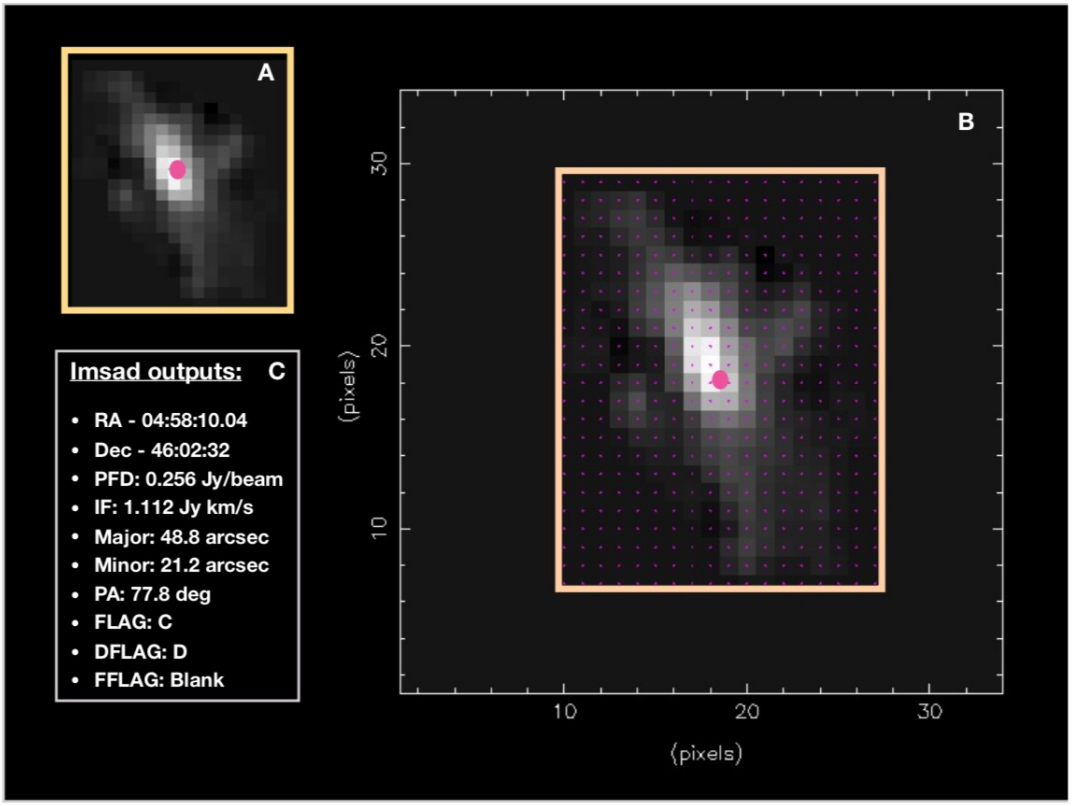


FIGURE 4.5: IMSAD Gaussian fit to a masked moment zero map. Panel A is a zoomed-out version of panel B, showing the HI line emission of the galaxy, the red dot indicates the derived flux centroid of the galaxy. Panel C shows source parameters measured IMSAD module.

Precision on galaxy coordinates

The recorded galaxy coordinates were identified from KVIS, visually. They give a good first approximation to the true flux centroid of the galaxy. However, if we desire precise measurements, more robust methods must be applied. This is because flux centroid of the detection might vary with respect to the channel viewed. Hence calculating flux centroid from the moment zero map is preferred in this thesis. This can be achieved by making use of the IMSAD module from MIRIAD package. IMSAD allows a user to derive the galaxy flux centroid by first fitting a Gaussian to the histogram of the moment zero map with an aim of determining the actual noise. Then identifies galaxies and fits Gaussians to each galaxy in order to calculate the centroids. Figure 4.5 shows an example of IMSAD when applied to galaxy candidate named J045811.74 + 460226.5. In addition, IMSAD can also calculate the

integrated flux by summing up all the flux of the galaxy. This is a slightly different way of calculating S_{int} than the one preferred in this thesis but both methods give consistent results.

4.1.4 Uncertainty measurements

The *precision*[1] and *accuracy*[2] of the derived HI parameters is of great importance for further analysis of the HIPP-ZoA data cube. However, these parameters inherited margins of error or uncertainties. There are three general types of errors that can occur when conducting an experiment: random errors; systematic errors and gross errors. Gross errors are due to human errors or equipment failure and are easily spotted outliers on the data and should be flagged out. Random errors are due to uncontrollable fluctuations in various variables or parameters that affect experimental end results. Whereas systematic errors are uncertainties inherited from instrument imperfections and methodological biases, which lead to lopsided results. Systematic errors can be corrected for in theory but require a full understanding of the entire system, hence in this project I won't calculate this type of uncertainty. On the other hand, random errors can be modelled using statistical methods and that is what I have done in this project.

Out of the total of nine derived HI parameters, five have random errors associated with them which will be determined. The errors associated with peak flux ($u(S_{\text{peak}})$) are calculated using an equation by Barnes et al. [2001]:

$$u(S_{\text{peak}}) = \sqrt{(\sigma_{\text{rms}})^2 + \left(\frac{S_{\text{peak}}}{20}\right)^2} \tag{4.9}$$

I then used equations from Koribalski et al. [2004] to calculated errors associated with integrated flux, $u(S_{\text{int}})$, and heliocentric velocity, $u(V_{\text{hel}})$:

$$u(S_{\text{int}}) = \left(\frac{4}{SNR}\right)\sqrt{(S_{\text{peak}} \times S_{\text{int}})\, dv} \tag{4.10}$$

$$u(V_{\text{hel}}) = \left(\frac{3}{SNR}\right)\sqrt{\left(\frac{W_{20} + W_{50}}{2}\right) dv}\,, \tag{4.11}$$

where dv is the channel width in km s^{-1} and SNR is the signal-to-noise ratio.

For errors associated with W_{20} & W_{50}, I used formulae by Schneider et al. [1986] , which associate the errors of these parameter with that of heliocentric velocity:

$$u(W_{20}) = 4u(V_{20}); \quad u(W_{50}) = 3u(V_{50}) \tag{4.12}$$

[1] Precision is the measure of consistency in a measurement when repeated.
[2] Accuracy is the measure of proximity to the "true" value.

Although these are not the only methods used in the literature to determine errors, the aforementioned formulae have been widely used by a large number of HI surveys (e.g. HIZOA: Staveley-Smith et al. [2016], ALFA-ZOA: McIntyre et al. [2015], ALFALFA: Giovanelli et al. [2005]), I hence adopt them for consistency. Secondly, the resulting measured uncertainties are consistent with those from the semi-automated catalogue.

4.2 Semi-automated source finding

The semi-automated source identification method is a blend between the visual and a fully automated method. This can be done by applying visual search on certain parts of the cube while automated source finding method is applied on other parts or apply automated search to the whole data cube then visually inspect the resulting data products (e.g. source masks, moment maps and global profiles) for detection confirmation. For this work, I used the outcome of a semi-automated method applied by Ramatsoku et al. [2016]. The semi-automated galaxy search was divided into two parts. The first part employed a customised script which uses various steps of the Groningen Image Processing SYstem (GIPSY: J. M. Van der Hulst and Roelfsema [1992]) to identify galaxy candidates. In part two, they had two individuals independently visualising all galaxy candidates from part one.
Before applying the customised script to the HIPP-ZoA data cube, they studied the noise characteristics across the 35 pointings and found variations (spatially). The variations were removed by dividing the data by the square of the weighted noise in each pointing (σ^2). They then applied a series of spatial and spectral smoothing to the HIPP-ZoA data cube before identifying galaxy candidates:

- Spatially smoothed the raw data cube (**RD**) of resolution $23'' \times 16''$ down to a beam size of $30'' \times 30''$ (**bs30**).
- Spectral smoothed **RD** and **bs30** to four spectral resolutions:
  1. Hanning smoothing (R2; 16.5 $\mathrm{km\,s^{-1}}$)
  2. Apply Gaussian kernel to 4 channels (R4; 33.0 $\mathrm{km\,s^{-1}}$)
  3. Apply Gaussian kernel to 6 channels (R6; 49.5 $\mathrm{km\,s^{-1}}$)
  4. Gaussian kernel applied to 8 channels (R8; 66.0 $\mathrm{km\,s^{-1}}$)

Figure4.6 shows a schematic with the eight data cubes. They applied the customised script on all eight cubes and identified 683 galaxy candidates. The detection was classified as a candidate if it had a flux density of at least: (i) 8σ in a single spectral resolution, (ii) 5σ in two adjacent spectral resolution, (iii) 4σ in three adjacent spectral resolution and (iv) 3σ in four adjacent spectral resolution. It must be noted that up to this point the source finding has been purely automated. After a systematic post visual inspection of all galaxy candidates, 235 out of 683 were identified as imaging artefacts or RFIs and were rejected.

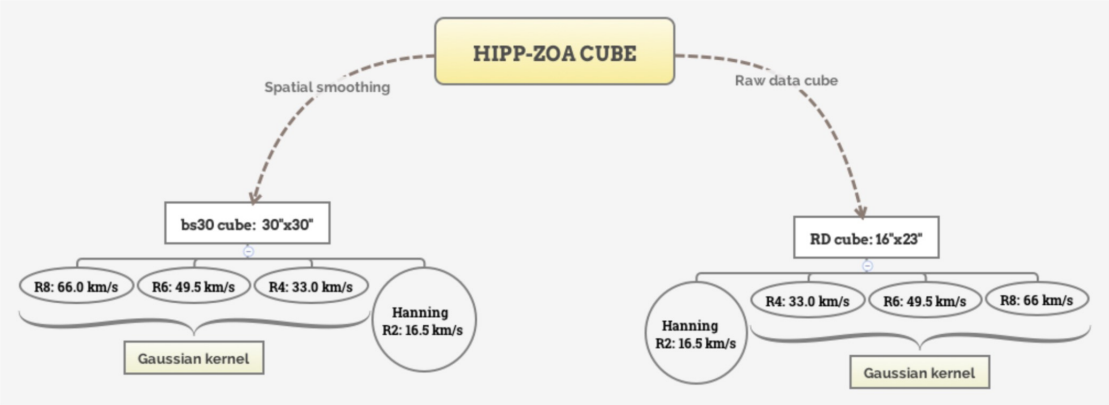


FIGURE 4.6: A sketch showing the eight-smoothed data cube of which the SA source finding method was applied to.

Further analysis led to the rejection of another 237 candidates as they had features consistent with noise peaks. This resulted in a semi-automated catalogue with 211 galaxies. A table that contains all the parameters associated with the 211 galaxies is presented by Ramatsoku et al. [2016]. In this thesis, we use their semi-automated catalogue to compare and contrast with that of visual and fully automated source finding method presented in the next section.

4.3 Fully-automated source finding

In Sect. 2.4.2 I have detailed steps and functions of the fully automated source finding application, SoFiA. This software was programmed to be applicable to HI datasets of different noise characteristics (e.g. datasets with uniform noise or with significant noise variations). The actual fine-tuning of SoFiA is heavily dependent on the noise properties of the dataset being inspected. SoFiA has had a handful of version since its release[3]. In this thesis I used SoFiA 1.2.0, which was the latest stable version (before submission of this thesis).

I fine-tuned SoFiA parameters based on the noise properties of the HiPP-ZoA HI data cube. After a number of practice runs of SoFiA to the data cube, I noticed the returned number of galaxies is extremely sensitive to just the slightest adjustments of the pipeline parameters. The optimal tuning of the pipeline is, therefore, a great challenge because it can either result in a large fraction of genuine galaxies being missed or return a large number of false positive.

To minimise such biases, I first identified (through visual inspection) a faint galaxy in the HiPP-ZoA data cube then extracted a subcube containing the galaxy. I then fine-tuned the SoFiA parameter file in such a way that it will identify the faint galaxy. I then extracted a subcube containing a massive spiral and tuned another SoFiA parameter file such that it identifies the massive spiral. The two tuned parameter files were then reconciled into

[3]https://github.com/SoFiA-Admin/SoFiA

one SoFiA parameter file. The reconciled parameter file should (in principle) be able to identify from faint to large massive galaxies.

A total of 64 out of 96 parameters (from the parameter file) were used. Out of the 64 parameters, nine parameters (i.e. *SCfinder.threshold, merge.radiusX, merge.radiusY, merge.radiusZ, merge.minSizeX, merge.minSizeY, merge.minSizeZ, reliability.threshold & reliability.fmin*) were fine-tuned several times to produce a more complete and reliable source catalogue. The listing below gives some of the most important parameter settings used to in arriving at this source catalogue (see: Table 5.3). The full parameter file is added under supplementary material (see: Appendix D).

The following steps were taken when applying SoFiA to the HiPP-ZoA data cube (remember the data cube is made up of nine subcube, see Fig. 3.1):

1. DATA PRECONDITIONING: Use external program (e.g. python) to convert *cdelt3* keyword value from $\mathrm{km\,s^{-1}}$ to $\mathrm{m\,s^{-1}}$.

2. Load data and sub-divide it into 9 subcubes with spatial dimensions of 750 : 750; adjacent subcubes have spatial and spectral overlap of 50 pixels and 50 channels, respectively.

3. FLAG BAD VOXELS: identified all regions with bad voxels and after a successful source finding run, all galaxy candidates that reside in these regions are flagged.

4. Apply a global noise normalisation filter.

5. SOURCE FINDING ALGORITHM: for this project I used the S+C algorithm with the following control parameters:

   - *SCfinder.threshold* = 3.0
     This is the flux detection threshold w.r.t HiPP-ZoA HI data cube.
   - *SCfind.kernels* = $[[0,0,0,'b'],[0,0,3,'b'],[0,0,7,'b'],[0,0,15,b],[3,3,0,'b'],[3,3,3,'b'],[3,3,7,'b'],[3,3,15,'b'],[6,6,0,'b'],[6,6,3,'b'],[6,6,7,'b'],[6,6,15,'b']]$
     These are 12 smoothing kernels used for the S+C source finder. The first two entries on each kernel represents number of pixels smoothed along the x-axis and y-axis, third entry is the number of channels smoothed and the fourth entry represents the smoothing kernels, where 'b' is the box-kernel and 'G' is the Gaussian-kernel.
   - *SCfind.rmsMode = negative*
     The root mean square noise (σ_{rms}) is derived by fitting a Gaussian to the negative flux of the data cube's histogram.

6. MERGING MODULE: This is one of the crucial steps of the pipeline as it defines the acceptable size of the genuine line emission source. It also informs the pipeline on how to cluster/group voxels into individual galaxy candidates and lastly, it flags out all the sources that do not meet the minimum requirements (in size) of a genuine source.

- *merge.radiusX = 3, merge.radiusY = 3* and *merge.radiusZ = 2*
  this defines the maximum distance (in pixels) in-between two neighbouring voxels that are part of one galaxy candidate.
- *merge.minSizeX = 5,merge.minSizeY = 5* and *merge.minSizeZ = 5*
  this is the minimum size (in pixels) of the genuine source in each axis after merging.

7. PARAMETRISATION MODULE:

   - *parameter.optimiseMask = true*
     chosen to use the mask optimisation algorithm base on fitting and growing ellipses for accurate flux measurements
   - *parameter.dilateMask = False*
     I decided against using the mask optimisation algorithm base on spatially dilating the initial mask

8. RELIABILITY MODULE:

   - *reliability.threshold = 0.95*
     discard all detections with reliability below this value.
   - *reliability.fmin = 10.0*
     is the minimum flux for a detection to be considered genuine
   - *reliability.makePlot = true*
     instructs the pipeline to produce reliability plots for visual purposes by the user. These plots are presented as supplementary material.

9. PIPELINE OUTPUT DATA PRODUCTS: For all the detected galaxy candidates, the pipeline produces a zeroth moment and a global profile.

   - *steps.doWriteMask = true*
     Produces a mask of the entire searched cube/subcube.
   - *steps.doWriteCat true*
     Writes a catalogue that contains HI parameters of all identified galaxy candidates (see Table B.1).
   - *steps.doMom0 = true*
     Produces a zeroth moment map for the entire data cube/subcube
   - *steps.doMom1 = true*
     Produces a first moment map for the entire data cube/subcube
   - *steps.doCubelets = true*
     If set to true, it produces the: mask, zeroth moment and first moment map for each galaxy candidate

After a successful run of SOFIA on all the nine subcubes making up the HIPP-ZOA HI data cube, a total of nine corresponding sub-catalogues were produced. All the sub-catalogues are then merged into one source catalogue which also consists of duplicate sources due

to the spatial overlap on the adjacent subcubes. If the derived HI were the same between the duplicates, one was flagged but if not, I inspected the duplicates maps and profiles then flagged the one that does not account for all the emission of the galaxy. The resulting catalogue is presented in the Table B.1.

Chapter 5

Results

In this chapter are the HI parameters for the two source catalogues derived in this thesis. The first catalogue (Visual) comprises of HI parameters for the 194 galaxy detections by the visual identification method. The second catalogue (SOFIA) named after the source finding method comprises of 131 galaxy detections.

5.1 The Visual catalogue

The Visual catalogue presented in Table A.1 consists of thirteen columns. The column entries of the 194 galaxies are as follows:

Column (1): A source designated identification number based in order of ascending right ascension.

Column (2): The galaxy identification name, derived from the coordinate system and sky coordinates. J marks J2000, and the next six digits are the ascension in hh:mm:ss, this is followed by the declination in dd:mm:ss.

Column (3): Indicates the subcube (SC) the galaxy was detected in.

Column (4) & **Column (5)**: Galactic longitude and Galactic latitude in degrees.

Column (6): The heliocentric velocity ($V_{\rm hel}$) of the galaxy in km s^{-1}.

Column (7) & **Column (8)**: HI profile linewidths in km s^{-1} measured at the 20% and level of the HI profile peak flux, respectively.

Column (9): The HI peak flux density $S_{\rm peak}$ in mJy. **Column (10)**: The integrated HI flux ($S_{\rm int}$) in Jy km s^{-1}.

Column (11): The cosmological distance (D) to the detected galaxy in Mpc and $H_0 = 70$ km s^{-1}/Mpc.

Column (12): The logarithm of the total HI mass in $M_\odot$.

Column (13): Flag indicating presence of a cross-match/counterpart galaxy from the semi-automated catalogue (SAC). "Y" means there exist a counterpart and "-" means no counterpart found.

TABLE 5.1: A sample of 22 galaxies identified from the WSRT HI data cube using visual identification method. The full catalogue is given in the appendix(Table. A.1).

No.	**Source ID**	**SC**	ℓ	b	$V_{\rm hel}$	W_{20}	W_{50}	$S_{\rm peak}$	$S_{\rm int}$	SNR	D	$\log(M_{\rm HI})$	SAC
			deg	deg	$\mathrm{km\,s^{-1}}$	$\mathrm{km\,s^{-1}}$	$\mathrm{km\,s^{-1}}$	$\mathrm{mJy\,beam^{-1}}$	$\mathrm{Jy\,km\,s^{-1}}$	$--$	Mpc	$\mathrm{M_\odot}$	$--$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	J044238.11 + 442618.1	9	160.11	−1.11	10696 ± 57	203 ± 71	171 ± 228	1.11 ± 0.40	0.04 ± 0.26	0.27	153	8.29	–
2	J044313.85 + 452731.7	3	159.41	−0.36	5480 ± 03	276 ± 09	245 ± 12	1.11 ± 0.40	0.60 ± 0.08	3.92	78	8.94	–
3	J044426.83 + 455117.4	3	159.25	+0.06	5487 ± 01	279 ± 03	265 ± 04	1.61 ± 0.41	1.07 ± 0.07	6.68	78	9.19	Y
4	J044522.55 + 454436.5	6	159.44	+0.11	5113 ± 02	298 ± 06	281 ± 08	2.88 ± 0.43	0.90 ± 0.11	5.42	73	9.05	Y
5	J044523.78 + 451929.3	6	159.76	−0.16	6811 ± 15	142 ± 45	63 ± 60	2.00 ± 0.41	0.21 ± 0.09	2.74	97	8.68	Y
6	J044540.46 + 454057.9	6	159.52	+0.11	9647 ± 07	145 ± 21	131 ± 28	2.04 ± 0.41	0.18 ± 0.14	1.62	138	8.91	Y
7	J044542.33 + 455043.6	3	159.40	+0.22	14564 ± 13	171 ± 39	71 ± 52	2.18 ± 0.41	0.27 ± 0.09	3.19	208	9.43	Y
8	J044544.03 + 442055.1	9	160.54	−0.75	10701 ± 04	198 ± 12	178 ± 16	2.37 ± 0.42	0.39 ± 0.12	2.98	153	9.33	Y
9	J044546.66 + 444858.0	6	160.19	−0.44	5128 ± 09	92 ± 27	55 ± 36	1.96 ± 0.41	0.23 ± 0.08	3.17	73	8.47	Y
10	J044550.65 + 443608.8	6	160.36	−0.57	16466 ± 05	88 ± 15	73 ± 20	1.12 ± 0.40	0.25 ± 0.07	2.93	235	9.51	Y
11	J044557.41 + 450758.1	6	159.97	−0.21	6785 ± 03	84 ± 09	70 ± 12	2.78 ± 0.42	0.48 ± 0.07	5.81	97	9.03	Y
12	J044559.21 + 450108.1	6	160.06	−0.28	12513 ± 12	148 ± 36	86 ± 48	1.45 ± 0.41	0.23 ± 0.08	2.51	179	9.24	Y
13	J044601.97 + 443442.8	6	160.40	−0.56	6265 ± 01	117 ± 03	99 ± 04	4.43 ± 0.46	2.42 ± 0.05	24.64	90	9.66	Y
14	J044628.17 + 451719.8	6	159.91	−0.04	10199 ± 16	137 ± 48	109 ± 64	1.75 ± 0.41	0.11 ± 0.15	1.10	146	8.75	Y
15	J044632.07 + 444313.6	6	160.35	−0.40	9846 ± 03	148 ± 09	131 ± 12	2.72 ± 0.42	0.42 ± 0.10	3.70	141	9.29	–
16	J044632.69 + 452156.9	6	159.86	+0.02	9722 ± 03	226 ± 09	209 ± 12	2.40 ± 0.42	0.53 ± 0.11	3.73	139	9.38	Y
17	J044635.70 + 441920.5	9	160.66	−0.65	9978 ± 07	132 ± 21	101 ± 28	2.37 ± 0.42	0.28 ± 0.10	2.79	143	9.12	Y
18	J044638.36 + 444632.6	6	160.32	−0.35	9893 ± 08	186 ± 24	45 ± 32	1.65 ± 0.41	0.54 ± 0.04	8.12	141	9.40	–
19	J044643.36 + 442030.5	9	160.66	−0.62	5637 ± 03	189 ± 09	171 ± 12	3.16 ± 0.43	0.55 ± 0.11	4.28	81	8.93	Y
20	J044646.08 + 444742.6	6	160.32	−0.32	5694 ± 03	161 ± 09	145 ± 12	2.85 ± 0.42	0.45 ± 0.11	3.83	81	8.85	Y
21	J044652.58 + 440918.6	9	160.82	−0.72	7488 ± 07	163 ± 21	138 ± 28	2.46 ± 0.42	0.26 ± 0.13	2.20	107	8.84	Y
22	J044655.23 + 444643.7	6	160.35	−0.31	16411 ± 11	121 ± 33	53 ± 44	1.19 ± 0.40	0.26 ± 0.06	3.58	234	9.52	Y

5.1.1 Discussion of the Visual HI parameters

I have presented the dataset and the derivation of HI parameters for identified galaxies. Here I show the distributions of the derived parameters from visual catalogues. Figure 5.1 presents six histograms of the HI parameters. Panel A shows the peak flux ($S_{\rm peak}$) histogram ranging from $\log(S_{\rm peak}/{\rm Jy}) = 0.04$ to 0.98. The distribution has an average of $log(S_{\rm peak}/{\rm Jy}) = 0.44$ and a spread of $log(S_{\rm peak}/{\rm Jy}) = 0.18$. A large fraction of 83.0% (161/194) of detections have peak flux greater than $\log(S_{\rm peak}/{\rm Jy}) = 0.20$ ($Speak = 1.58$ mJy beam^{-1}).

Panel **B** is the distribution of integrated flux. It covers a range of $S_{\rm int} = 0.03$ to 10.02 Jy km s^{-1} and peaks at 0.66 Jy km s^{-1}. A large fraction of fluxes (76.29%, 148/194) are confined within a smaller range of $\log(S_{\rm int}/{\rm Jy\,km\,s^{-1}}) = -0.90$ to 0.00 ($0.13 < S_{\rm int} < 1.00$ Jy km s^{-1}).

Panel **C** presents the W_{20} histogram. The distribution spans a range of 37 to 397 km s^{-1} and peaks at 164 km s^{-1}. The lopsided distribution is typical of the linewidth distribution found in literature (e.g. ALFALFA, Giovanelli et al. [2005]).

Panel **D** shows the W_{50} linewidth distribution, spanning a range of 21 to 378 km s^{-1}. The distribution has a mean linewidth of $W_{50} = 126$ km s^{-1}, which is less than that of σ.40 catalogue (Haynes et al. [2011]). However, this difference is not statistically significant due to the large scatter in the two distributions. A large fraction of 92.3% (179/194) detections have W_{50} measurements less than 250 km s^{-1}.

Panel **E** presents the histogram of signal-to-noise ratio (SNR). The distribution spans a range of $SNR = 0.27$ to 65.47 and has a mean of 5.92. The lopsided distributions show large number of detections have low signal-to-noise ratio.

Panel **F** presents the total HI mass distribution spanning range of $\log(M_{\rm HI}/{\rm M}\odot) = 7.8$ to 10.2. This is less in comparison to the most massive HI galaxies in the literature, which can reach HI masses of 10.8 (e.g. Cluver et al. [2010]). The distribution has an average HI mass of $\log(M_{\rm HI}/{\rm M}\odot) = 9.1$, which is less than $\log(M_{\rm HI}/{\rm M}\odot) = 9.5$ of ALFALFA (Giovanelli et al. [2005]). About $\sim$ 10% of detections have HI masses below $\log(M_{\rm HI}/{\rm M}\odot) = 8.5$.

A summary of the HI parameters distributions is contained in Table. 5.2. The first column contains HI parameters, column 2 is the spanned range, column 3 present the mean of the distribution the last column has the standard deviation.

TABLE 5.2: A summary of the distribution of measured HI parameters from Visual.

Panel	**HI parameters**	**Range**	**Mean**	**std** (σ)
A	$S_{\rm peak}$ [Jy beam^{-1}]	1.11 – 9.6	2.76	1.52
B	$S_{\rm int}$ [Jy km s^{-1}]	0.03 – 10.02	0.67	1.00
C	W_{20} [km s^{-1}]	37 – 397	164	75
D	W_{50} [km s^{-1}]	21 – 378	126	75
E	SNR	0.27 – 65.47	5.92	7.37
F	$\log(M_{\rm HI}/{\rm M}_\odot)$	7.81 – 10.24	9.08	0.46

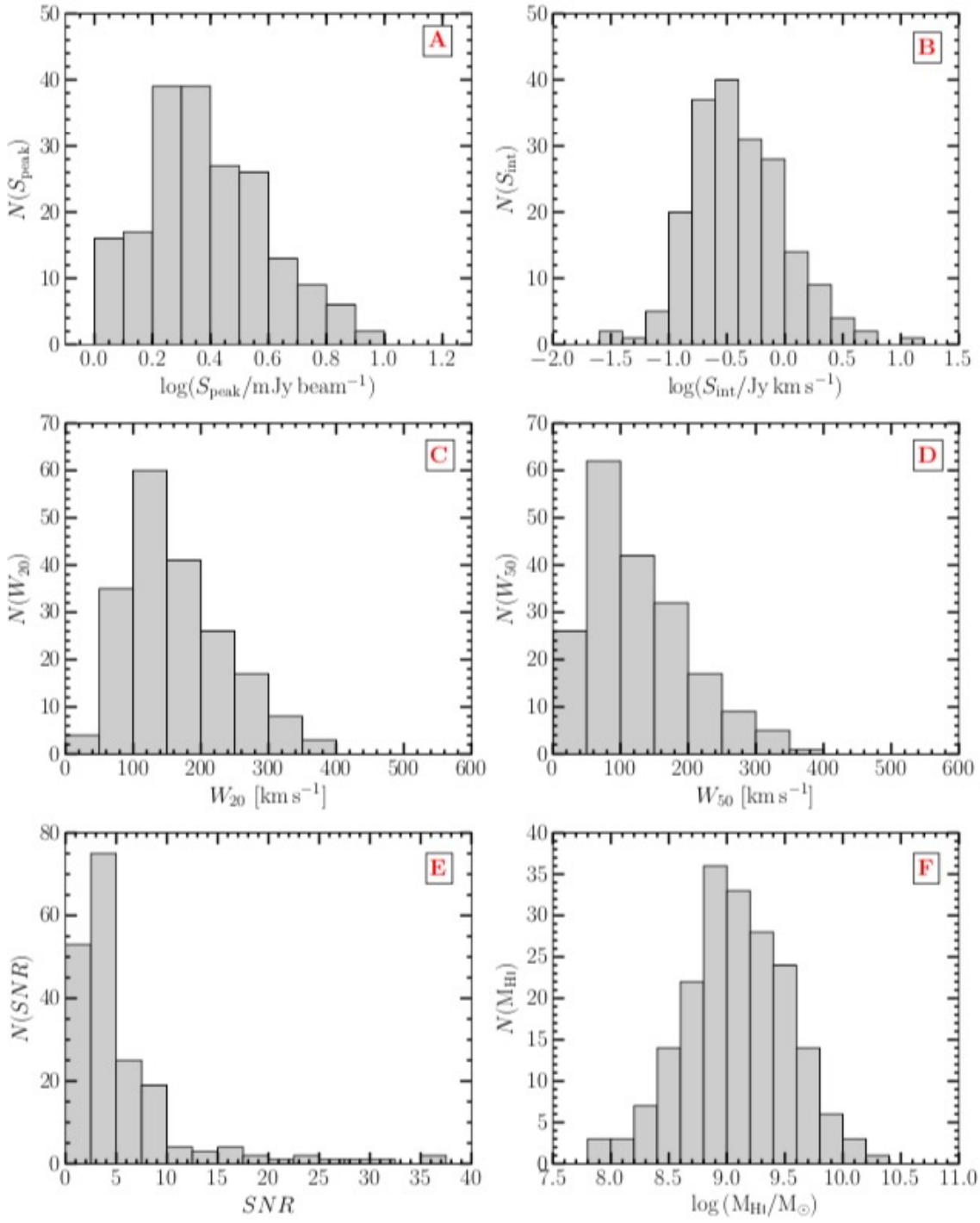


FIGURE 5.1: The distribution of measured HI parameters of visual detections. **A**: the logarithmic of the peak flux ($S_{\rm peak}$). **B**: the logarithmic of $S_{\rm int}$. **C**: the linewidth at 20% of peak flux W_{20}. **D**: the linewidth at 50% of peak flux W_{50}. **E**: signal-to-noise ratio (SNR). **F**: the logarithm of total HI mass).

TABLE 5.3: A sample of the first 22 galaxy detections by a fully automated source finder, SOFIA. The full catalogue is given in the appendix (Table. B.1).

No.	Source ID	SC	ℓ	b	$V_{\rm hel}$	W_{20}	W_{50}	$S_{\rm peak}$	$S_{\rm int}$	SNR	D	Mass	CM
			deg	deg	$\rm km\,s^{-1}$	$\rm km\,s^{-1}$	$\rm km\,s^{-1}$	$\rm mJy\,beam^{-1}$	$\rm Jy\,km\,s^{-1}$	$--$	Mpc	$\rm M_\odot$	$--$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	J044521.33 + 454433.8	6	159.44	+0.11	5099 ± 01	294 ± 03	277 ± 04	3.20 ± 0.43	0.93 ± 0.11	5.64	73	9.06	Y
2	J044542.93 + 442103.0	9	160.54	−0.75	10687 ± 08	432 ± 24	170 ± 32	2.40 ± 0.42	0.30 ± 0.13	2.36	153	9.22	Y
3	J044556.86 + 450740.7	6	159.97	−0.21	6776 ± 01	86 ± 03	70 ± 04	2.80 ± 0.42	0.34 ± 0.09	4.15	97	8.88	Y
4	J044602.28 + 443415.4	9	160.41	−0.56	6252 ± 00	110 ± 00	97 ± 00	4.60 ± 0.46	2.10 ± 0.05	21.65	89	9.60	Y
5	J044632.73 + 452147.5	6	159.86	+0.02	9697 ± 03	332 ± 09	250 ± 12	3.60 ± 0.44	1.04 ± 0.10	6.65	139	9.67	Y
6	J044635.79 + 441908.8	9	160.66	−0.65	9968 ± 01	126 ± 03	97 ± 04	2.40 ± 0.42	0.34 ± 0.09	3.53	142	9.21	Y
7	J044644.14 + 442002.3	9	160.67	−0.62	5619 ± 01	193 ± 03	175 ± 04	3.60 ± 0.44	0.66 ± 0.11	5.08	80	9.00	Y
8	J044644.75 + 444735.4	6	160.32	−0.32	5687 ± 01	154 ± 03	137 ± 04	2.90 ± 0.43	0.48 ± 0.10	4.16	81	8.87	Y
9	J044652.85 + 440900.6	9	160.82	−0.72	7471 ± 02	156 ± 06	112 ± 08	2.50 ± 0.42	0.32 ± 0.11	3.09	107	8.94	Y
10	J044700.42 + 442441.1	9	160.64	−0.54	10710 ± 03	520 ± 09	309 ± 12	5.10 ± 0.47	0.85 ± 0.15	4.89	153	9.67	Y
11	J050256.91 + 451232.1	4	161.80	+2.16	3243 ± 00	145 ± 00	124 ± 00	5.80 ± 0.49	1.43 ± 0.08	13.07	46	8.86	Y
12	J044600.22 + 442444.1	9	160.52	−0.67	13017 ± 06	90 ± 18	66 ± 24	6.20 ± 0.51	0.08 ± 0.25	1.03	186	8.83	–
13	J044602.52 + 443427.9	6	160.40	−0.56	6247 ± 00	112 ± 00	97 ± 00	4.60 ± 0.46	2.69 ± 0.05	27.77	89	9.70	–
14	J044736.90 + 444754.1	6	160.41	−0.20	9756 ± 07	133 ± 21	55 ± 28	1.70 ± 0.41	0.13 ± 0.10	1.77	139	8.77	–
15	J044817.40 + 445346.4	6	160.42	−0.05	13947 ± 23	151 ± 69	117 ± 92	1.40 ± 0.41	0.04 ± 0.23	0.36	199	8.55	–
16	J044819.56 + 444647.3	6	160.51	−0.12	16507 ± 04	65 ± 12	36 ± 16	1.50 ± 0.41	0.11 ± 0.08	1.87	236	9.16	–
17	J044841.48 + 455316.0	3	159.70	+0.64	12370 ± 15	1191 ± 477	374 ± 636	1.90 ± 0.41	0.08 ± 0.34	0.43	177	8.78	–
18	J044845.25 + 441450.9	9	160.97	−0.40	6606 ± 16	324 ± 48	$144. \pm 64$	1.20 ± 0.40	0.12 ± 0.14	1.02	94	8.40	–
19	J044902.56 + 450145.4	6	160.40	+0.14	12726 ± 23	635 ± 69	375 ± 92	2.40 ± 0.42	0.19 ± 0.25	0.99	182	9.17	–
20	J044902.60 + 450145.6	6	160.40	+0.14	11936 ± 02	144 ± 06	93 ± 08	2.50 ± 0.42	0.68 ± 0.07	7.16	171	9.67	–
21	J044943.90 + 460245.3	3	159.70	+0.88	12743 ± 04	656 ± 12	251 ± 16	8.20 ± 0.57	0.65 ± 0.20	4.15	182	9.70	–
22	J044945.22 + 460234.0	3	159.70	+0.88	9466 ± 01	123 ± 03	81 ± 04	18.70 ± 1.02	0.43 ± 0.21	4.88	135	9.27	–

5.2 SoFiA catalogue

The SoFiA catalogue (Table 5.2 & B.1) has the same structure as the visual catalogue except for the last column. The columns are as follows:
Columns (1) - (13): are the same as the corresponding columns in visual catalogue.
Column (14): Y: Indicates that the galaxy has a crossmatch in the Visual catalogue, whereas - means there is no crossmatch.

5.2.1 Distributions of the SoFiA HI parameters

In the previous section, I presented histograms of HI parameters from the visual catalogue. These histograms are also produced for the SoFiA detections to allow a comparison of the distributions and understand the differences between two, for instance, what kind (e.g. in linewidth parameter space) of galaxies are missed by one method compared to the other.

The histograms are presented in Figure 5.2. The top left panel, **G** is the Speak histogram, H present the $S_{\rm int}$ histogram, I is the W_{20} while **J** contains the W_{50} histogram, **K** present the SNR distribution and **L** shows the logarithmic total HI mass distribution. These distributions are also presented (numerically) in table below.

TABLE 5.4: A summary of the distribution of measured HI parameters from SoFiA.

Panel	HI parameters	Range	Mean	std (σ)
G	$S_{\rm peak}$ [Jy beam^{-1}]	$0.80 - 18.70$	3.82	2.45
H	$S_{\rm int}$ [Jy km s^{-1}]	$5.67 \times 10^{-3} - 9.77$	0.75	1.07
I	W_{20} [km s^{-1}]	$29 - 656$	175	103
J	W_{50} [km s^{-1}]	$11 - 375$	123	74
K	SNR	$0.14 - 66.97$	6.80	8.31
L	$\log(M_{\rm HI}/{\rm M}_\odot)$	$7.71 - 10.08$	9.09	0.46

The SoFiA $S_{\rm int}$ histogram shows that SoFiA identified more faint galaxy candidates than Visual. However most of these detections could be false positive and are discussed in the next chapter. The W_{20} histogram has a range of $29 - 656$ km s^{-1} which is wider compared to the Visual range of $37 - 397$ km s^{-1}. It must be noted that only 5.67% detections have W_{20} greater than 400 km s^{-1} . The W_{50} distribution spans a range of $11 - 375$ km s^{-1}, while the Visual range is $21 - 378$ km s^{-1}. A small fraction of 4.25% SoFiA detections measured have the W_{50} less 21 km s^{-1} and they all do not have cross-matches in other catalogues. The histogram of the SNR spans a range of $0.07 - 75.87$ km s^{-1}, whereas the Visual range is $0.27 - 65.47$ km s^{-1}. It must be noted that all the SoFiA detections with signal-to-noise greater than $SNR = 10.0$ have Visual crossmatches. The SoFiA detections span an overall mass range of $\log(M_{\rm HI}/{\rm M}_\odot) = 7.71$ to 10.08 and an average of $\log(M_{\rm HI}/{\rm M}_\odot) = 9.09$.

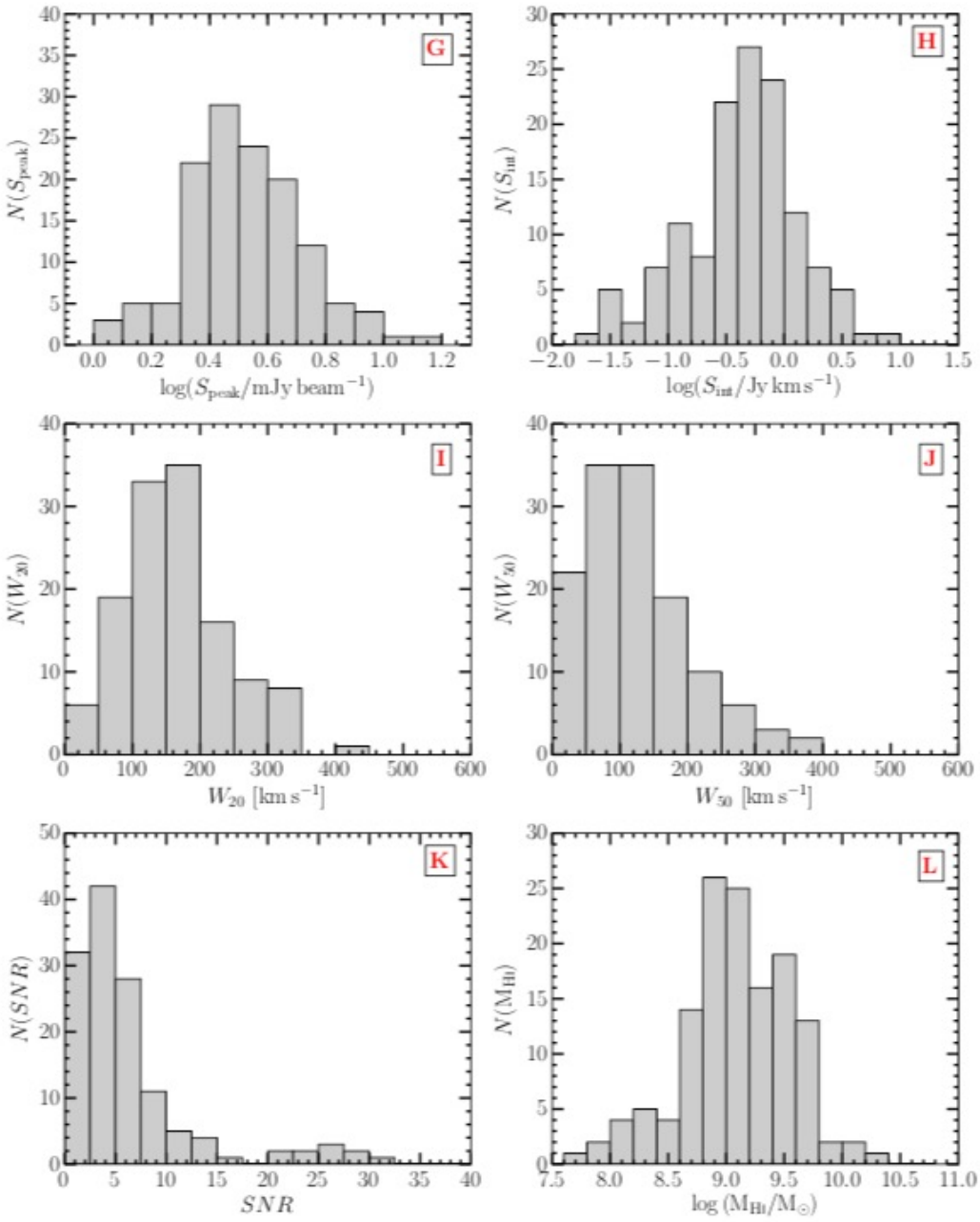


FIGURE 5.2: The distribution of measured HI parameters based on SOFIA. **G**: the logarithmic of the peak flux (S_{peak}). **H**: the logarithmic of S_{int}. **I**: the linewidth at 20% of peak flux W_{20}. **J**: the linewidth at 50% of peak flux W_{50}. **K**: signal-to-noise ratio (SNR). **L**: the logarithm of total HI mass).

5.3 Sensitivity curves

The sensitivity curve is a distribution of the HI mass as a function of the heliocentric velocity of all detected galaxies. It shows where detected galaxies of certain linewidths and HI masses lie with respect to the detection threshold. All sensitivity curves presented in this work, assume two 3σ detection thresholds, namely $\mathcal{L}_{3\sigma,100}$ and $\mathcal{L}_{3\sigma,200}$. The first detection threshold assumes a linewidth of $W_{50} = 100$ km s^{-1}, while the latter assumes a linewidth of $W_{50} = 200$ km s^{-1}. The galaxy detections are grouped into the five linewidths brackets (i.e. Ψ_i) base on their respective 50% linewidth measurements.

$$\Psi_i = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \\ \Psi_5 \end{bmatrix} = \begin{bmatrix} \Psi_1 \leq 50 \text{ km s}^{-1} \\ 50 < \Psi_2 \leq 100 \text{ km s}^{-1} \\ 100 < \Psi_3 \leq 200 \text{ km s}^{-1} \\ 200 < \Psi_4 \leq 300 \text{ km s}^{-1} \\ \Psi_1 > 300 \text{ km s}^{-1} \end{bmatrix}$$

5.3.1 Visual

Figure 5.3 consists of three sub-plots illustrating the sensitivity curve, heliocentric velocity and HI mass distributions for all Visual detections. The main panel (bottom left) shows the sensitivity curve. A fraction of 6.7 (13/194) galaxies lie below the first 3σ detection threshold ($\mathcal{L}_{3\sigma,100}$) and their measured linewidths are below $W_{50} = 200$ km s^{-1}. A further 25.8% (50/194) detections lie in-between the two detection thresholds (i.e. $\mathcal{L}_{3\sigma,100}$ and *LsigmaTwo*). The remaining 67.5% are found above the $\mathcal{L}_{3\sigma,200}$ detection threshold and the measured linewidths span the all the linewidths brackets. Only a fraction of 5.7% (11/194) of the detections have HI masses greater than the characteristic mass of HI mass function, $\log(M^*_{\rm HI}) = 9.8$ [Zwaan et al., 2005], compared to the 8.1% (17/211) of Semi-automated detections. It must be noted that the data cube has increased noise at heliocentric velocities above $V_{\rm hel} = 15\,500$ km s^{-1} and only one galaxy has been visually identified in this region.

The bottom left panel illustrates the total HI mass distribution. The histogram shows a fairly symmetric distribution about the mean mass of 1.20×109 M$_\odot$. From the symmetry we can deduce that the ratio of low mass galaxies to high mass is approximately equal to one. This is quite typical distribution for a schechter type HI mass function with fixed sensitivity. The distributions also shows that galaxies with HI masses less than $\log(M_{\rm HI}/{\rm M}_\odot) = 9.5$, have linewidths less than $W_{50} = 300$ km s^{-1}.

The top panel present a histogram of the heliocentric velocity. The histogram covers a range of $V_{\rm hel} = 3\,213$ to $16\,466$ km s^{-1} and has an average of $V_{\rm hel} = 8562$ km s^{-1}. The distribution shows regions of overdensities, which were first identified by Ramatsoku et al. [2016]. These overdensities are found on the Auriga (Aur) constellation. With the nearest overdensity (Aur 1) at a velocity of $3\,000 - 3\,500$ km s^{-1} and consisting of 9 galaxy detections, whereas the second overdensity (Aur 2) spans a velocity range of

$4500 - 7500$ km s^{-1} and has a total of 81 detections. The third overdensity (Aur 3) is enclosed over a range of $9500 - 11000$ km s^{-1} and has the second highest number of detections (68). The last overdensity (Aur 4) has 20 detections and is spread over a narrow velocity range of $13\,500 - 15\,500$ km s^{-1}. In-between the overdensities, we have undersense regions with a significantly low number of galaxies compared to overdensities, see also Fig. 18 of Ramatsoku et al. [2016].

To gauge how the HI masses and other measured HI parameters of the Visual detections compare to the other galaxy identification methods, I will present a comprehensive comparison in Ch. 6.

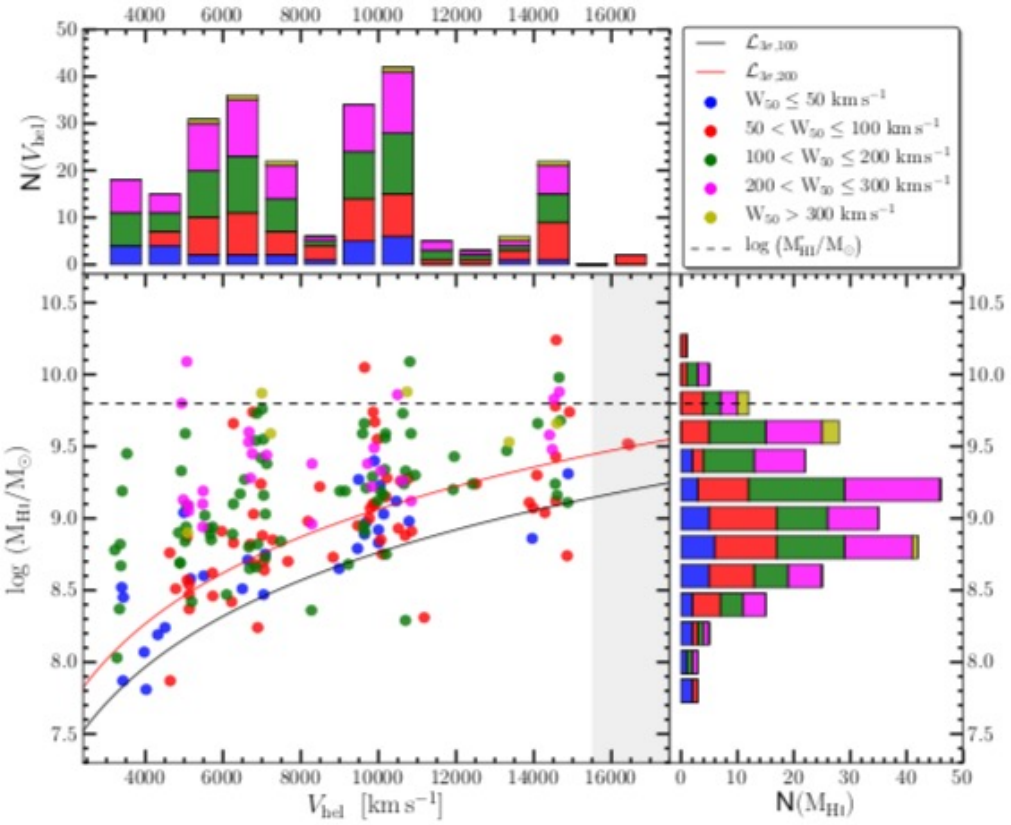


FIGURE 5.3: The efficiency of the Visual method presented in terms of the sensitivity curve (bottom left panel), the distribution of the logarithmic HI mass (bottom right panel) and the heliocentric velocity histogram (top left panel).

5.3.2 SoFiA

Figure 5.4 illustrates the sensitivity curve, HI mass and heliocentric velocity histograms of all 131 SoFiA detections. The bottom left panel present the sensitivity curve. About 16.2% (21/131) detections lie below the $\mathcal{L}_{3\sigma,100}$ detection threshold. A small fraction of 8.% (11/131) galaxies are bounded by the two detection thresholds ($\mathcal{L}_{3\sigma,100}$ and $\mathcal{L}_{3\sigma,200}$). This is fraction is three times less than that of Visual detections. This shows that current

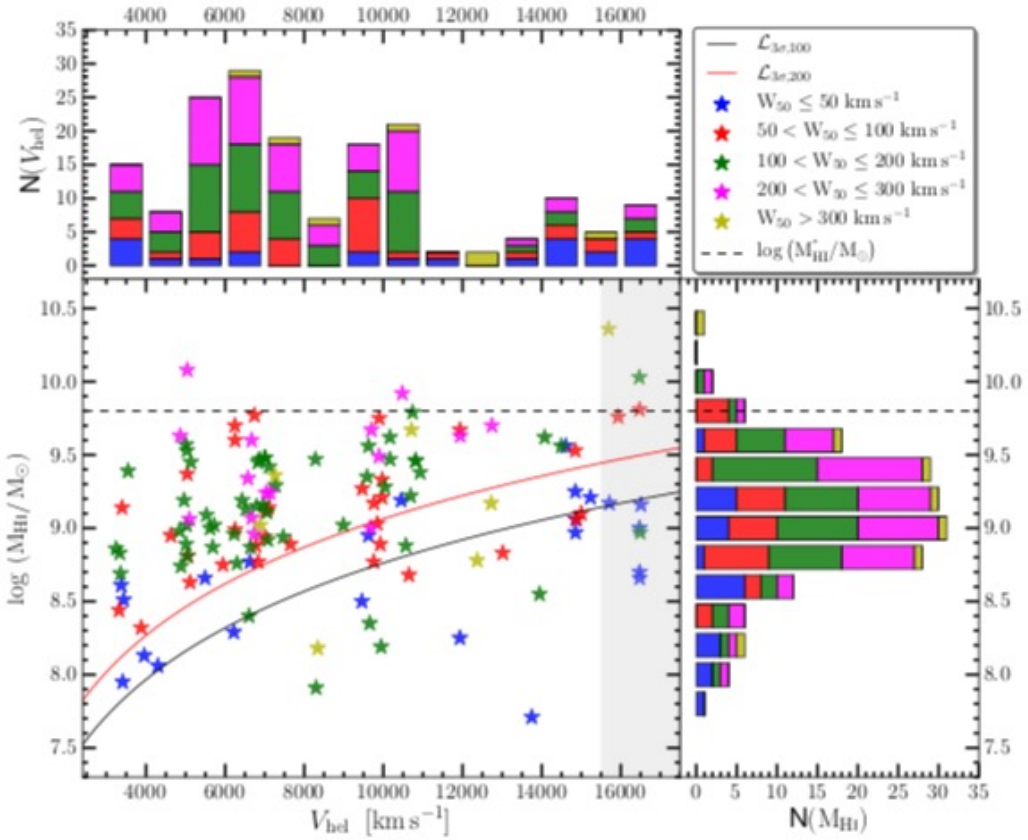


FIGURE 5.4: I present three sub-figures illustrating the radial velocity distribution; the sensitivity curve and the total HI mass distribution of all 131 galaxies identified through SOFIA.

parameter setup of SOFIA performs poorly in this region. A further fraction of 75.3% of detections lie above the $\mathcal{L}_{3\sigma,200}$ detection threshold and cover all the linewidths brackets.

The bottom right panel shows the HI mass distribution. This distribution is also presented in Fig. 5.2, however in this panel I want to illustrate the distribution with respect to linewidth brackets (Ψ_i). Note that galaxies with linewidths in the first bracket (Ψ_1), have HI masses less than $\log(M_{HI}/M_{\odot}) = 9.5$. This is consistent with the Visual HI mass distribution presented in Fig.5.1. Only 3.1% of the detection have HI masses greater than the $\log(M^{*}_{HI})$ whereas Visual has 5.7%.

The heliocentric velocity distribution is shown in the top panel. It covers a wide velocity range of $V_{hel} = 3\,243$ to $16\,507$ km s^{-1}. However the mean radial velocity of the distribution is $V_{hel} = 8678$ km s^{-1}. The overdensities seen in Fig.5.3 are also prominent in this panel. The first overdensity (Aur 1) consists of 8 detections, whereas Aur 2 has 56, Aur 3 is the second largest overdensity with 28 detections and Aur 4 is made up 12. Recall, Visual method identified one galaxy candidates with heliocentric velocity greater than $5\,500$ km s^{-1}, however SOFIA has identified 10. This rise is due to the high rms noise in this region and this is a phenomenon known as edge effects and it affects all automated source fining tools.

This chapter was dedicated to the parameters and properties of the two catalogues, Visual and SoFiA. Visual has a total of 194 galaxy detections and 32.5% found below the 3σ detection limit assuming a 200 km s^{-1} linewidth ($\mathcal{L}_{3\sigma,200}$). Only one detection was made at the region with poor signal-to-noise ratio (i.e. $V_{\rm hel} \geq 15\,500$ km s^{-1}). The second catalogue, SoFiA, revealed 131 detections. SoFiA has identified $\sim$ 8% less detections under the $\mathcal{L}_{3\sigma,200}$ detection threshold compared to Visual. About 10 SoFiA detections have radial velocities that are within the poor signal-to-noise ratio. An in-depth comparison and analysis of these catalogues and semi-automated catalogue is presented in the next chapter.

Chapter 6

Discussion

In this section, I present a detailed analysis of the Visual, Semi-automated and SoFiA catalogue. In Ch. 4, I presented the three different source identification algorithms that were applied to arrive at a source catalogue and the techniques that were employed to derive coordinates and HI parameters. All these techniques were applied to the same HiPP-ZoA data cube. This provides, therefore, an ideal Testbed to check the consistency of all HI parameters in all 3 catalogues as well as a comparison of their respective completeness. This is pursued by taking the following steps:

- Identify counterparts/cross-match galaxies between Visual and Semi-automated catalogue, Visual and SoFiA catalogue and SoFiA and Semi-automated catalogue.
- Comparison of HI parameters amongst the counterparts' galaxies.
- Employ sensitivity plots to illustrate where the galaxy counterparts are found with respect to redshift and also the two HI mass limit detection curves, $\mathcal{L}_{3\sigma,100}$ and $\mathcal{L}_{3\sigma,200}$.

Identifying counterparts and analysing the consistency of the HI parameters is not sufficient to know how each of these galaxy identification methods performs with respect to the completeness or missed galaxies in each catalogue and how these methods can be improved and optimised. To get to that point, will need to understand characteristics of the subsets of detections that have no counterpart in the other catalogues. Where do they lie on sensitivity plots? What are their typical properties in linewidths, peak flux, SNR, HI mass, etc.? Is there any possible optimisation necessary to improve both completeness and reliability achieved with the various method? These are the questions I intend to tackle in this chapter.

6.1 Galaxy counterparts identification

Galaxies have two spatial Galactic coordinates (ℓ & b) and the third coordinate is the radial velocity ($V_{\rm hel}$). These are the coordinates that I have used to identify counterparts. Supposedly galaxy **A** has coordinates (ℓ^A, b^A, V^A) and that galaxy **B** is the counterpart. That means that the coordinates of the counterpart must satisfy the relations below:

$$(\ell, b, V_{\rm hel})^B \Longleftrightarrow \begin{bmatrix} (\ell^A - \Delta S) \leq \ell^B \leq (\ell^A + \Delta S) \\ (b^A - \Delta S) \leq b^B \leq (b^A + \Delta S) \\ ({V_{\rm hel}}^A - \Delta V) \leq {V_{\rm hel}}^B \leq ({V_{\rm hel}}^A + \Delta V) \end{bmatrix}$$

where ΔS is a spatial radius enclosing the surface area where a counterpart galaxy is most likely to be found, whereas ΔV is the radius along the velocity axis.

A graphical presentation of the above equation is presented in Fig. 6.1. This schematic illustrates concentric circles of radii known as counterpart clouds, where a probability of finding a true counterpart is inversely proportional to the radii. The spatial radii are a function of the pixel size from the original HI data cube. The pixel size is equivalent to $pixel - size = 6''$, hence the spatial radii of the counterpart clouds are: $r = 3(pixel - size)$, $r' = 5(pixel - size)$, $r'' = 7(pixel - size)$. Whereas the velocity radii are: $r_v = 100$ km s^{-1}, $r'_v = 150$ km s^{-1} , $r''_v = 200$ km s^{-1}. Using this schematic, crossmatch candidates are identified. Afterwards, moment maps of the galaxy and its possible counterpart(s) are visualised to qualify or disqualify the counterpart. The latter is particularly important because the different techniques might subdivide galaxies with satellites or very extended galaxies in different counterparts.

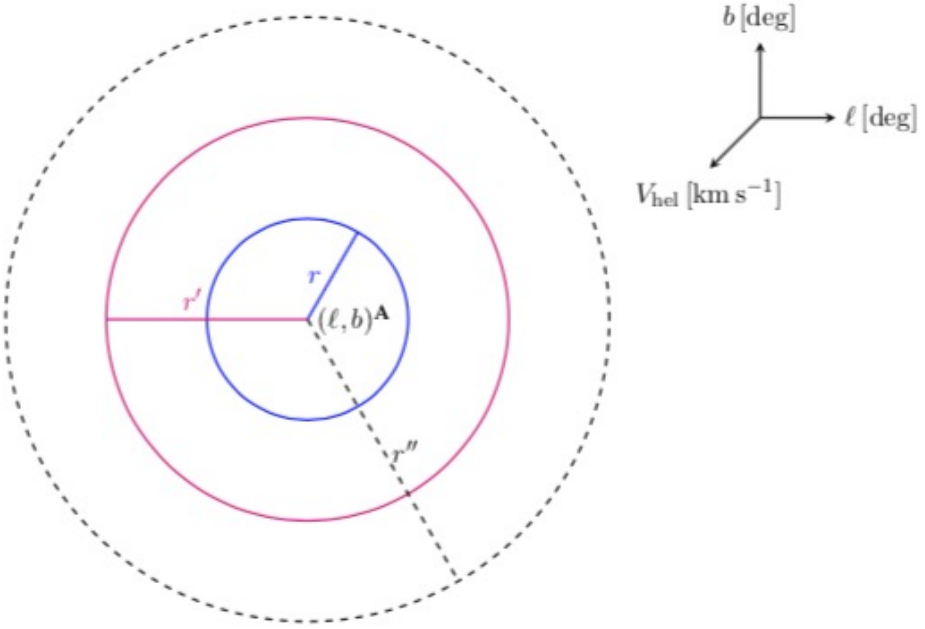


FIGURE 6.1: A projection of concentric spherical counterpart clouds (spheres) centred at galaxy **A** onto 2-D spatial plane. The inner circle marks a cloud with radius of $r = 3(pixel - size)$, red circle has $r' = 5(pixel - size)$ and the black circle has $r'' = 7(pixel - size)$.

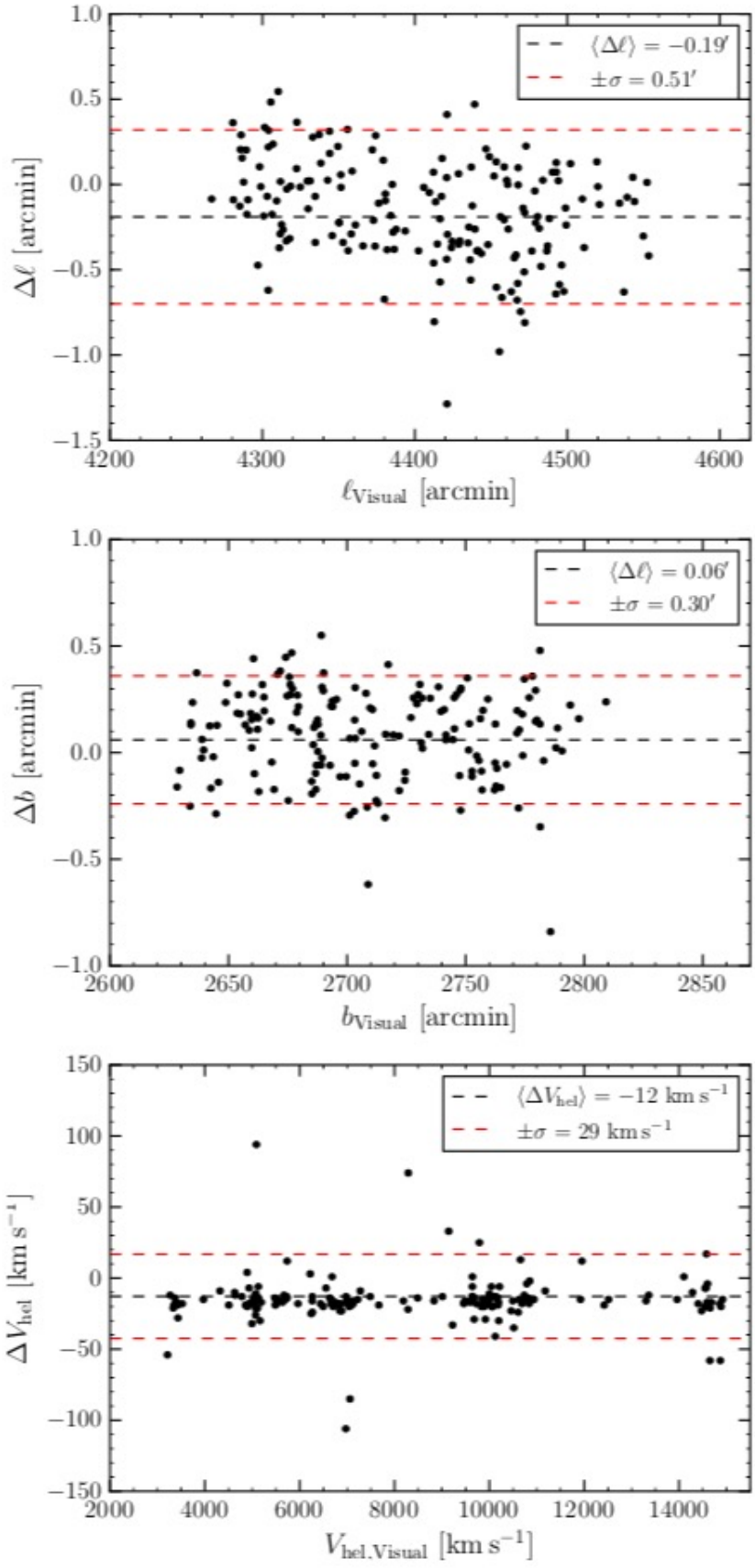


FIGURE 6.2: A comparison of coordinates of Visual galaxies against their Semi-automated crossmatches. The top panel: $\Delta\ell$ vs ℓ. Middle panel: Δb vs b. Bottom panel: $\Delta V_{\rm hel}$ vs $V_{\rm hel}$.

6.2 Visual and Semi-automated

The visual catalogue presented in Table 5.1 consists of 194 galaxies. Using the aforementioned galaxy counterpart identification criteria, a total of 89.7% (174/194) galaxies, have counterparts in the Semi-automated galaxy catalogue published by Ramatsoku et al. [2016]. In this first section, I will focus on comparing the HI parameters of these Visual galaxies against their counterparts (V-SA counterparts).

6.2.1 Comparing measured HI parameters

Figure 6.2 presents dispersion plots between the Visual coordinates and the corresponding Semi-automated counterparts. The top panel illustrates the Galactic longitude dispersion centred at an average of $\Delta\ell = -0.19$ arcminutes and has a standard deviation of $\sigma = 0.51$ arcminutes. This dispersion plot spans a range of $\Delta\ell = -4.18$ to 1.40 arcminutes.

The Galactic latitude dispersion plot is presented in the middle panel. The distribution spans a range of $\Delta b = -2.64$ to 0.55 arcminutes and has an average of $\Delta b = 0.06$ arcminutes and a standard deviation of $\sigma = 0.30$ arcminutes.

The bottom panel compares heliocentric velocities of the crossmatches. The distribution spans a range of $\Delta V_{\rm hel} = -106$ to 217 km s^{-1} and is centred at an average of $\left\langle \Delta V_{\rm hel} \right\rangle = -12$ km s^{-1}, with a standard deviation of $\sigma = 29$ km s^{-1}.

The coordinates of the Visual galaxies are consistent with the coordinates of their respective Semi-automated crossmatches. The identified crossmatches are within a maximum spatial displacement of $\Delta S = 0.58$ arcminutes (0.01 deg). The velocity dispersion shows a spread of 29.6 km s^{-1} but has a broader range. This is typical of HI galaxy velocity comparisons. This is due to HI sources being extended along spectral axis. For instance a deviation of 5 pixel translate to a dispersion of 41.3 km s^{-1}.

The next step is comparing the measured HI parameters. This is achieved by first identifying outliers (mathematically) then explore why such dispersions occur. Presented in Fig. 6.3 are four dispersion/scatter plots. Where the dots present the data, dashed black line marks mean and the red dashed lines present the plus/minus 3 sigma level. The top left panel illustrates the scatter between the 20% linewidth of Visual galaxies and their Semi-automated counterparts. The $W_{20}-$dispersion spans a range of $\Delta W_{20} = -97$ to 217 km s^{-1} and is centred at 2.6 km s^{-1} with a standard deviation of $\sigma = 34.2$ km s^{-1}. There are four galaxies with W_{50} greater than 3-sigma. The W_{50}-dispersion plot is presented on the top right panel. It spans a range of -214 to 169 km s^{-1}. The distribution has an average of -4.5 km s^{-1} and a standard deviation of 36.5 km s^{-1}. Both the W_{20} and W_{50}-dispersion show Gaussian distributions with similar mean and standard deviation measurements. The $W_{50}-$dispersion plot reveals seven counterparts with large scatter in the 50% linewidth measurements and two also have scatter in W_{50}. We will look into some of the global profiles of these sources later on.

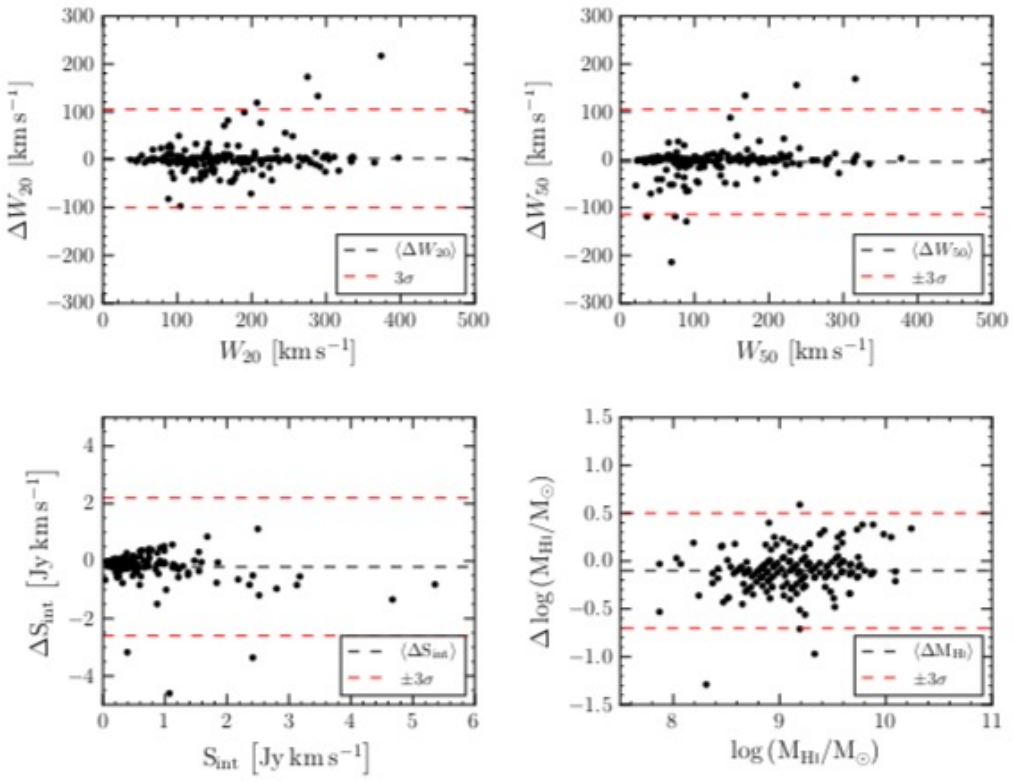


FIGURE 6.3: Comparison of HI parameters between a total of 89.7% (174/194) galaxies from visual catalogue (along x-axis) against galaxy counterparts (174/211) from semi-automated catalogues (along y-axis).

The bottom left panel, shows the $S_{\rm int}$-dispersion plot. It spans a range of -7.61 to 1.12 Jy km s^{-1} and has a mean of -0.2 Jy km s^{-1} with an rms of 0.80 Jy km s^{-1}. About 93% dispersion measurement are within one sigma and only four have deviations of more than 3-sigma, however all four show small deviations in W_{20} and W_{50}.

The dispersions observed in integrated flux and heliocentric velocity plots feed to the total HI mass dispersion plot shown in the bottom right panel. The dispersions are centred at an average of $\Delta \log(M_{\rm HI}/{\rm M}_\odot) = -0.10$ and returns an rms of 0.23. The dispersions reveal four counterparts with deviations in the logarithm of HI mass greater than 3-sigma. Two of these counterparts also show larger deviations in integrated flux while only one show larger dispersions in the 50% linewidths. This is a common phenomenon with methods that incorporate automated source finders as then turn to merge satellites with their parent sources or fragment faint galaxies with broad profile galaxies. Will discuss this in the next section where we compare SOFIA with both Visual and SA source catalogues.

6.2.2 The relative completeness

Out of the 194 Visual galaxies, 20 were not identified by the Semi-automated source identification method. On the other hand, 37/211 Semi-automated galaxies have no Visual crossmatches. In this section, will address the relative completeness of the two catalogues

as a function of linewidth, heliocentric velocity, signal-to-noise ratio and total HI mass.

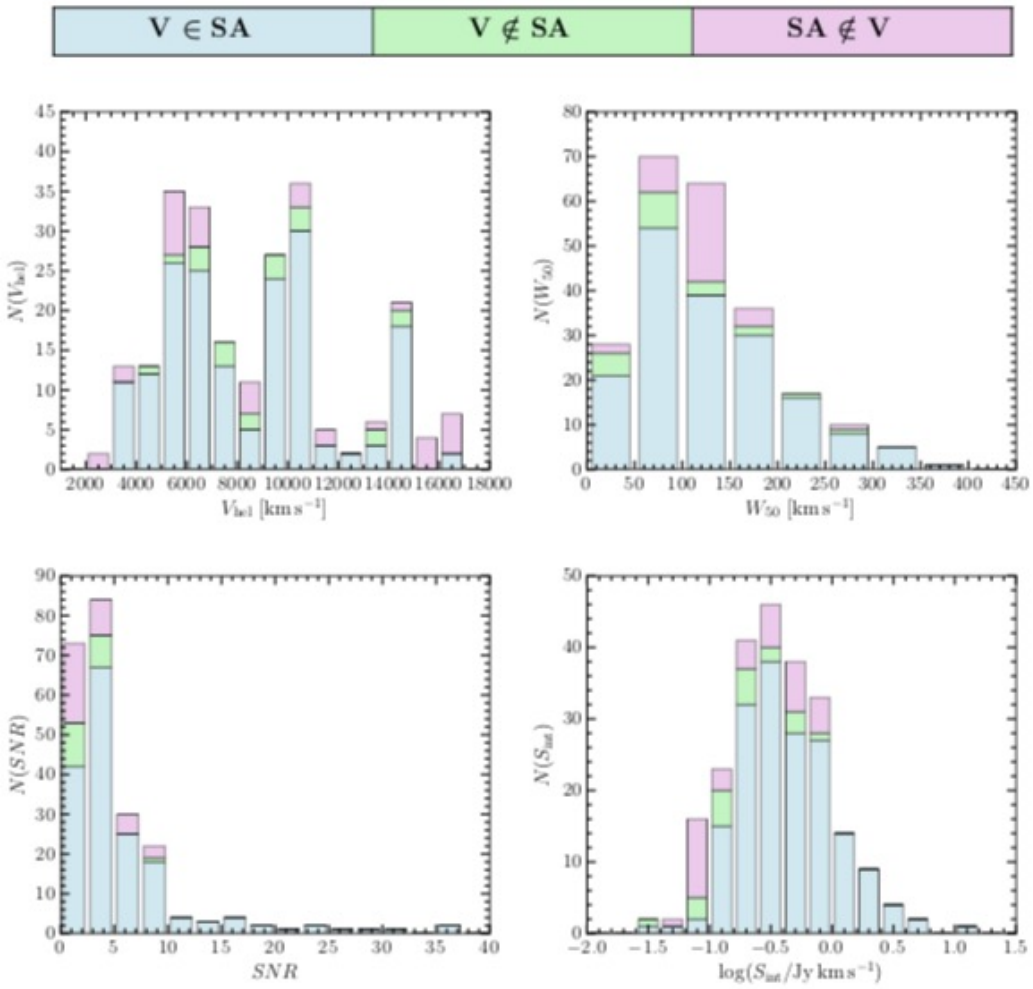


FIGURE 6.4: Distributions of HI parameters for both common and non-common galaxies between Visual and Semi-automated. Top left panel: W_{50} stacked histogram. Top right panel: V_{hel} histogram, bottom left panel: signal-to-noise ratio and bottom right panel: the logarithm of HI mass distribution.

Figure 6.4 present histograms which illustrate the relative completeness of Visual and Semi- automated catalogue. Marked in light-blue are distributions of Visual galaxies with no SA crossmatches, while SA with no Visual crossmatches are shown in green and crossmatches are shown in grey. Shown in the top left panel is the velocity histogram. The crossmatches span a range of $3\,213 - 16\,466$ km s^{-1} with an average of $8\,483$ km s^{-1}. All the 9 Visual detections under Aur. 1 have SA crossmatches, whereas a fraction of 74/81 Aur. 2. Under Aur. 3, 48/68 Visual galaxies have SA crossmatches and 19/20 in the fourth overdensity (Aur. 4). At the high-end velocity ($V_{hel} > 15\,500$ km s^{-1}), the 2 Visual galaxies have SA crossmatches. However, there are 8/10 SA detections with no Visual crossmatches. Due to poor signal-to-noise within this region, we do not have sufficient information to classify these detections as false positives or true detections. SA has 13 detection in Aur. 2 that have no crossmatches compared to 7 Visual detections. This shows that in

regions of higher overdensities and poorer signal-to-noise, semi-automated tools are more likely to return a high number of false positives.

The 50% linewidth distribution is shown in the top right panel. The crossmatches distribution spans a range of $21 - 378$ km s^{-1} and is centred at 129 km s^{-1}. The two catalogues are complete above $W_{50} = 300$ km s^{-1} and there are only two Visual detections with no SA crossmatches within the range of $200 - 300$ km s^{-1} compared to one SA detection. About 59% of SA detections with no Visual crossmatches have W_{50} (22/37) between $100 - 150$ km s^{-1}, whereas $\sim$ 65% (13/20) of Visuals with no SA crossmatches are below $W_{50} = 100$ km s^{-1}.

On the bottom left panel, is the signal-to-noise ratio histogram. All the Visual detections with the signal-to-noise ratio greater than $SNR = 10.0$ have SA crossmatches and vice versa. Only 1 out of 44 Visual detections with SNR in-between the range of $5 - 10$ does not have SA crossmatch, whereas 8 out of 50 SA detections have no Visual crossmatches. Below the signal-to-noise ration of $SNR = 5$, Visual identified 19 galaxies with no SA crossmatches while SA method has identified 29 without Visual crossmatches. The visual method has achieved a completeness (w.r.t SA method) of 98.5% (65/66) above the signal-to-noise ratio of $SNR = 5$ and 85.2% (109/128) below this threshold.

The bottom right panel illustrates the distribution of the logarithm of integrated flux. The histogram covers a range of $\log\left(S_{\rm int}/\text{Jy km s}^{-1}\right) = -1.47$ to 1.00 and is centred at -0.36. Both catalogues agree on all the detections with integrated flux great than $\log\left(S_{\rm int}/\text{Jy km s}^{-1}\right) = 0.00$ $\left(S_{\rm int} = 1.00 \text{ Jy km s}^{-1}\right)$. About 80% (16/20) of the Visual detections with no SA crossmatches have integrated fluxes that lie within the range of $\log\left(S_{\rm int}/\text{Jy km s}^{-1}\right) = -1.00$ to 0.00 $(0.10 < S_{\rm int} < 1.00\ S_{\rm int})$, compared to the $\sim$ 68% (25/37) SA detections with no Visual crossmatches. There are 20% (4/20) Visual detections with $\log\left(S_{\rm int}/\text{Jy km s}^{-1}\right)$ less than -1.00 which have no SA crossmatches whereas SA has identified $\sim$ 32% (12/36). SA has identified $1.5\times$ more sources with integrated flux less than $S_{\rm int} < 0.10$ Jy km s^{-1} as compared to Visual.

All the parameters presented in Fig. 6.4 determine where a detection lies in the sensitivity plot. Figure 6.5 present the sensitivity plot of the Visual detections with and without SA crossmatches. The colours in both panel are classified as follows: i) blue present Visual galaxies with SA crossmatches (V $\in$ SA), ii) green is Visual with no SA crossmatches (V $\notin$ SA) and iii) pink illustrates SA with no Visual crossmatches (SA $\notin$ V).

The right panel shows the distribution of the logarithm HI mass. The crossmatches span a range of $\log\left(M_{\rm HI}/\text{M}_\odot\right) = 7.9$ to 10.24 and an average of $\log\left(M_{\rm HI}/\text{M}_\odot\right) = 9.1$. The Visual galaxies with no SA crossmatches are distributed over a mass range of $\log\left(M_{\rm HI}/\text{M}_\odot\right) = 7.8$ to 9.8 and are centred at 8.8 with an rms of 0.46. SA detections with no Visual span the range of $7.7 - 9.9$ with an average of 8.7 and a narrow rms of 0.55. Above $\text{M}^*_{\rm HI}$, Visual has only one detection with no crossmatch compared to two of SA. About 65% of Visual detections without SA crossmatches have HI masses below $\log\left(M_{\rm HI}/\text{M}_\odot\right) = 9.0$ compared to the $\sim$ 60% SA detections with no Visual crossmatches.

The sensitivity plot on the left panel is sub-divided into three regions (R I, R II, and

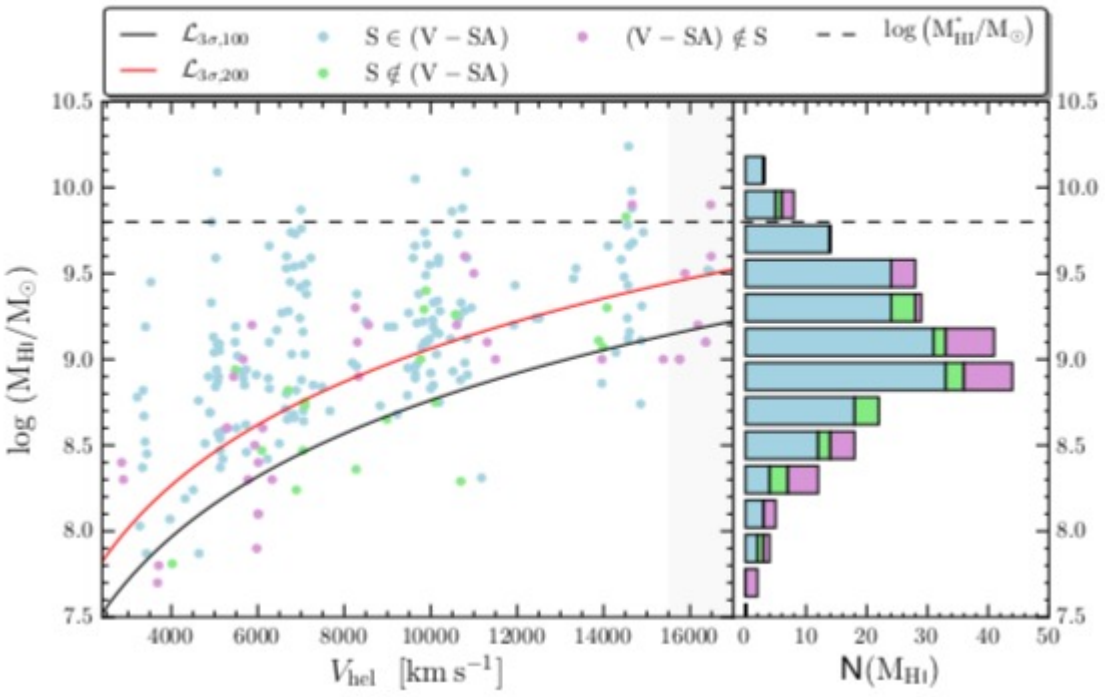


FIGURE 6.5: The distribution of total HI mass as a function heliocentric velocity. Shown in blue are the 174 Visual galaxies with SA crossmatches, whereas green marks the 20 Visual detections with no SA crossmatches and in pink are the 37 SA sources with no visual crossmatches. The left panel shows the sensitivity curve while right panel shows histogram of the total HI mass.

R III,). R I is the region above the 3σ detection threshold assuming a linewidth of 200 km s^{-1} ($P_i > \mathcal{L}_{3\sigma,200}$). R II is in-between the $\mathcal{L}_{3\sigma,200}$ and the $\mathcal{L}_{3\sigma,100}$ (i.e. $\mathcal{L}_{3\sigma,100} < P_i < \mathcal{L}_{3\sigma,200}$) and R III lies below the $\mathcal{L}_{3\sigma,100}$. Let us look at the relative completeness and reliability of these method based on the aforementioned regions. At R I, about 95.5% Visual galaxies have SA crossmatches. While in R II the two catalogues have a large fraction of 83.7% common detections. Then in R III Visual has identified 13 galaxies and 7 of them were also detected by SA. Overall Visual has 89.7% (174) detections that have SA crossmatches.

The summary of the relative efficiency between these two catalogues is presented in Table 6.1. Where the first column indicates regions of the sensitivity plot. Column 2 contains Visual detections and is sub-divided detections with crossmatches and with none, they are marked as TD and FP, respectively. Third column is the total number of SA detections.

TABLE 6.1: Efficiency of the Visual method with respect to the Semi-automated method (SA).

	Visual		**SA**	**Reliability**	**Completeness**
	TD	FP	TS	$\mathfrak{R}$ [%]	$\mathfrak{C}$ [%]
RI	126	6	146	95.5	86.3
RII	41	8	49	83.7	83.7
RIII	7	6	16	53.8	43.8
Survey	**174**	**20**	**211**	**89.7**	**82.5**

Column 3 is the relative reliability given by:

$$\mathfrak{R} = \frac{\mathrm{TD}}{\mathrm{TD} + \mathrm{FP}} \tag{6.1}$$

Column 5 is the relative completeness and can be derived using the equation below:

$$\mathfrak{C} = \frac{\mathrm{TD}}{\mathrm{TD} + \mathrm{TS}} \tag{6.2}$$

The two galaxy identification methods identified a large number of common galaxies (174) and have measured fairly consistent HI parameters. This is a great result because now we have a sufficient sample to compare and contrast with a catalogue output of the fully automated software, SOFIA. The Testbed sample is key in analysing SOFIA results as it offers a way of quantifying relative reliability and completeness.

6.3 Testbed sample and SOFIA

The Testbed sample contains 174 galaxies that have been identified by both the Visual and SA catalogue. It must be noted that the actual number of galaxies in the HIPP-ZoA data cube is unknown. This is typical of spectral cubes derived from blind surveys. Hence, quantifying the efficiency of an automated source-finder such as SOFIA can be a challenge. For this work, we assume that the Testbed sample represents the actual number of galaxies in the HIPP-ZoA data cube.

This section has a layout similar to Sect. 6.1. I will first identify common galaxies between the SOFIA and the Testbed catalogue, then assess the consistency of position and HI parameters. Finally, will quantify reliability and completeness of the SOFIA catalogue when assuming Testbed sample.

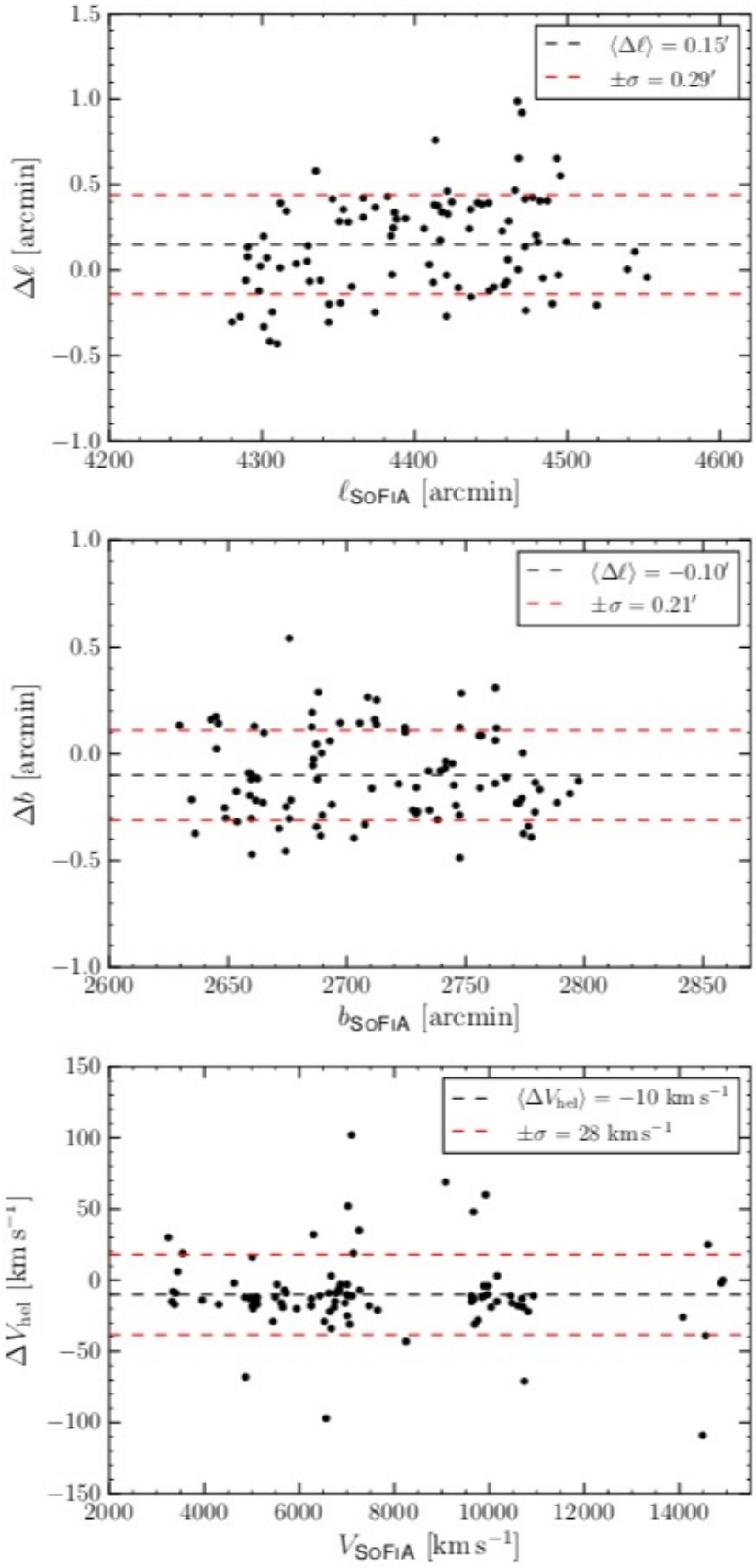


FIGURE 6.6: A comparison of coordinates between the S-TB crossmatches. The top panel: $\Delta\ell$ vs ℓ. Middle panel: Δb vs b. Bottom panel: ΔV_{hel} vs V_{hel}.

6.3.1 Consistency of positions and HI parameters

Using the crossmatch identification scheme in Sect. 6.1 a total of 91 out of 131 SoFiA detections are also listed in the Testbed catalogue.

The positional consistency of these crossmatches is presented in Fig. 6.6. The top left panel presents the the Galactic longitude dispersion plot. The distribution covers a range of $\Delta\ell - 0.43$ to 0.99 arcminutes and is centered at 0.15 arcminutes with an rms of 0.29 arcminutes. The Galactic latitude dispersion plot is shown in the middle panel. It spans a range of $\Delta b - 0.49$ to 0.55 arcminutes with an average and an rms of -0.10 and 0.21 arcminutes, respectively. The velocity dispersion plot is centered at $\left\langle \Delta V_{\rm hel} \right\rangle = -10$ km s^{-1} and has an rms of 28 km s^{-1}. It covers a range of $V_{\rm hel} - 109$ to 102 km s^{-1}.

The positional dispersion encountered here are less than those observed between the Visual and the SA crossmatches (Sect.6.2.1). Hence, we can conclude that the source positions returned by SoFiA are consistent with those measured by the Visual and SA methods.

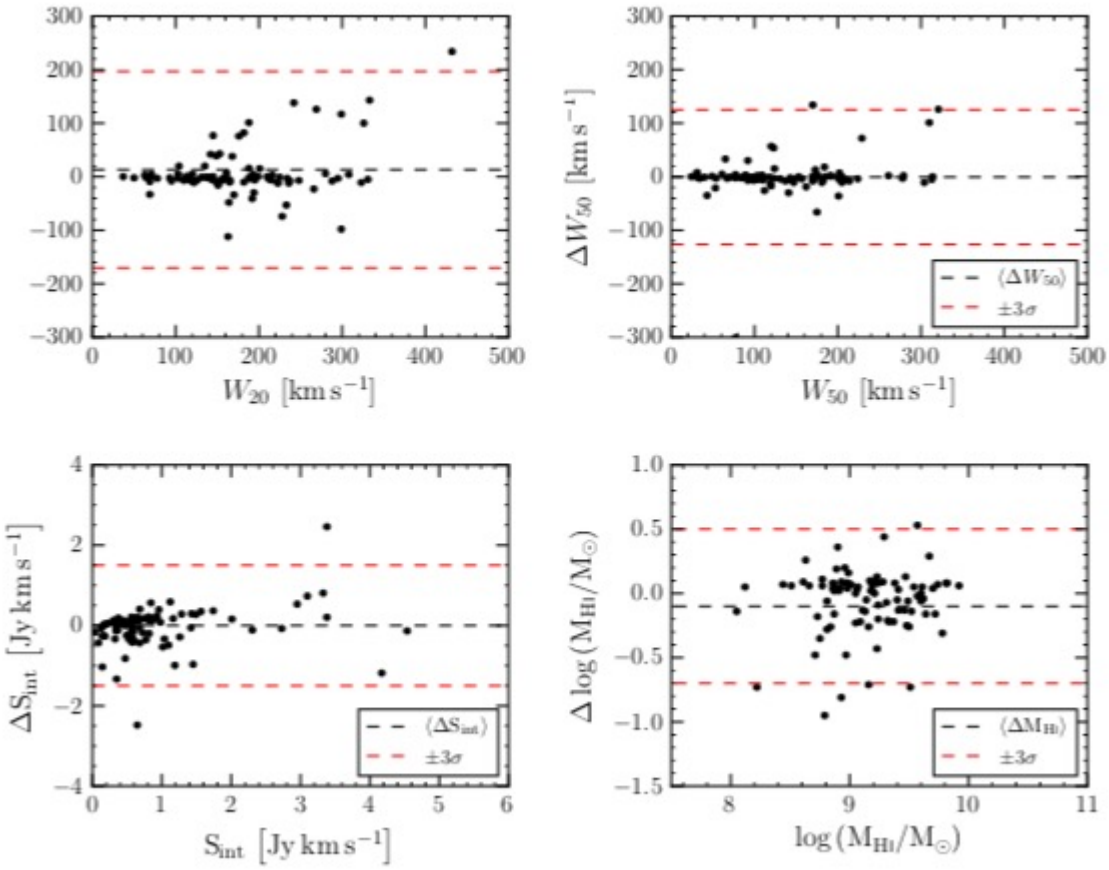


FIGURE 6.7: Comparison of HI parameters between a total of 89.7% (174/194) galaxies from the visual catalogue (along x-axis) against galaxy counterparts from (174/211) from semi-automated catalogues (along y-axis).

Figure 6.7 presents four HI parameter dispersion plots. The top left panel shows the 20% linewidth dispersion plot, centered at $\Delta W_{20} = 13$ km s^{-1} and has an rms of 61 km s^{-1}. About 3.3% (3/91) of crossmatches show dispersions greater $3\times$ rms. Will look at these sources later on, to see why such deviations occur.

The top right panel shows the 50% linewidth dispersions, spanning a range of $\Delta W_{20} = -301$ to 134 km s^{-1}. The distribution has an average of 0 km s^{-1} and an rms of 42 km s^{-1}. A total of 96.7% (88/91) show consistent measurements. One of the three outliers also showed greater deviations in W_{20} dispersion plot.

The integrated flux dispersion plot is presented in the bottom left panel. The distribution covers a range of $\Delta S_{\rm int} = -2.5$ to 2.5 Jy km s^{-1} and is centred at zero with an rms of 0.5 Jy km s^{-1}. The dispersions observed here are fairly consistent and only 2 sources show dispersion larger deviation. These sources also reveal larger deviations in W_{50} measurements.

The bottom right panel illustrates the logarithmic HI mass dispersions. It spans the range of $\Delta \log(M_{\rm HI}/{\rm M}_\odot) = -0.95$ to 0.53 and has an average of $\Delta \log(M_{\rm HI}/{\rm M}_\odot) = -0.06$ with an rms of 0.25. Five sources show larger (>3 rms) deviations and one of them also show larger deviations in $V_{\rm hel}$, W_{50} and $S_{\rm int}$. Comparing just the numerical parameters is not enough, as there are a number of scenarios that can lead to these variations.

Hence moment maps of crossmatches with inconsistent measures were examined. Four examples of crossmatches with slightly inconsistent HI parameters are presented in Fig. 6.8 to Fig. 6.11. Figure 6.8 shows a scenario where SSOFIA has successful identified two tidal interacting galaxies, however the integrated flux measured by SOFIA is slightly lower and results to a lesser HI mass. Figure 6.9 present a galaxy with low SNR and has a profile buried in the noise. SOFIA successfully identified this galaxy but also missed faint emission at the edges. In Fig. 6.10 is a detection obscured by a negative peak, hence such dispersions in measured parameters. Figure 6.11 shows a case where only one horn of the galaxy has been detected and this leads to underestimating linewidths, integrated flux and in turn the total HI mass.

The majority of SOFIA detections show consistency in derived positions and HI parameters when compared to their crossmatches from the Testbed parameters. There a very few cases were SOFIA underestimates or overestimate certain HI. The next step is quantifying the efficiency of the pipeline.

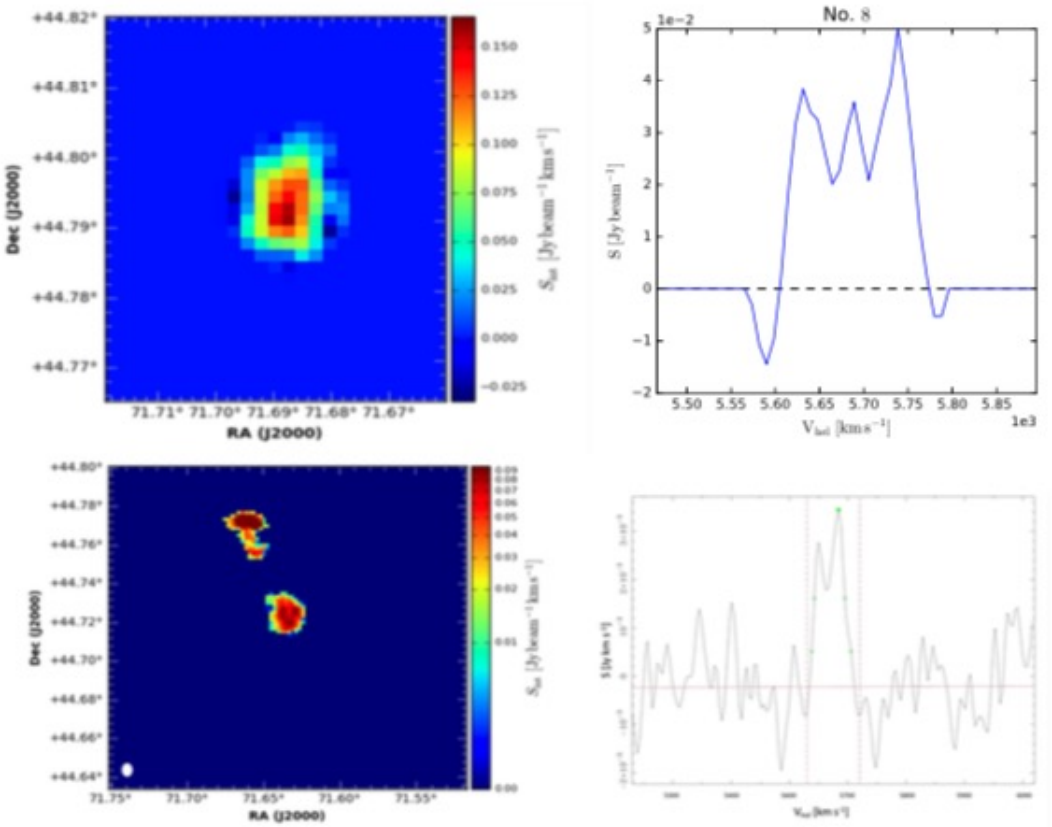


FIGURE 6.8: A SOFIA catalogue with a corresponding crossmatch from the Testbed sample. Top panels: Maps of galaxy no. 8 from SOFIA. Bottom panels: Maps of galaxy no. 1 from Visual..

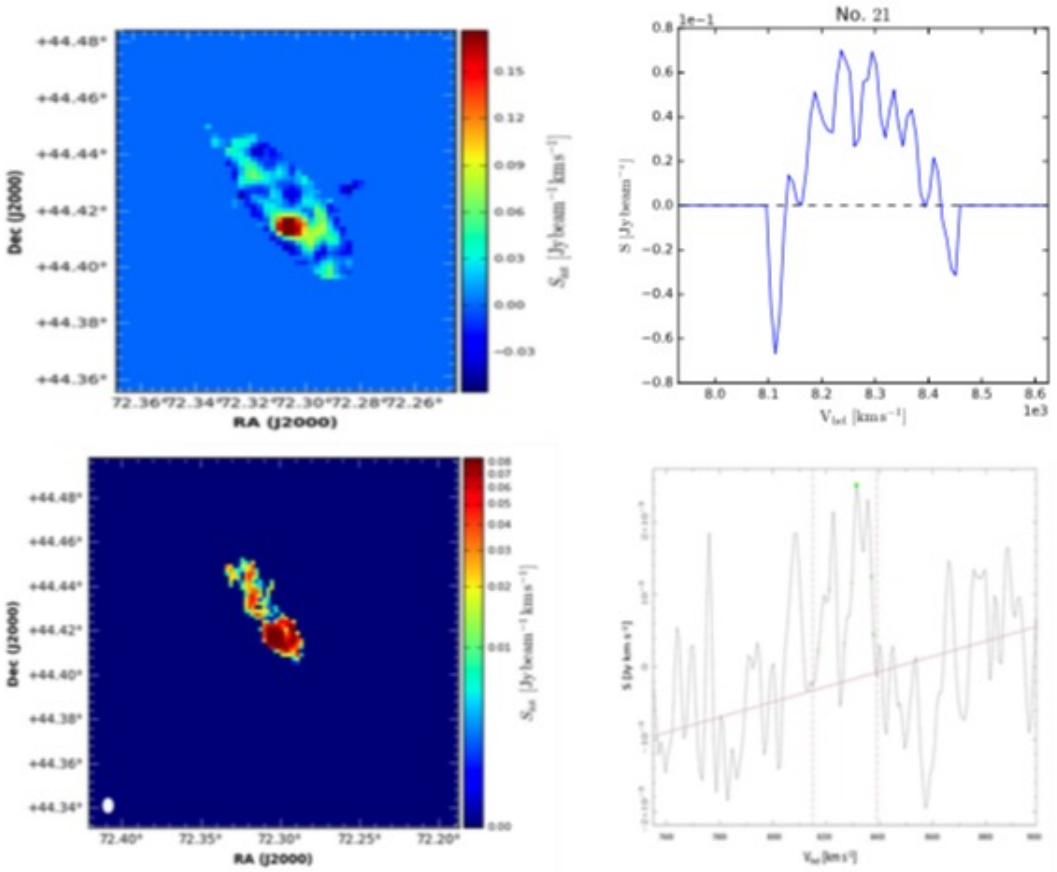


FIGURE 6.9: A SOFIA catalogue with a corresponding crossmatch from the Testbed sample. Top panels: Maps of galaxy no. 21 from SOFIA. Bottom panels: Maps of galaxy no. 2 from Visual.

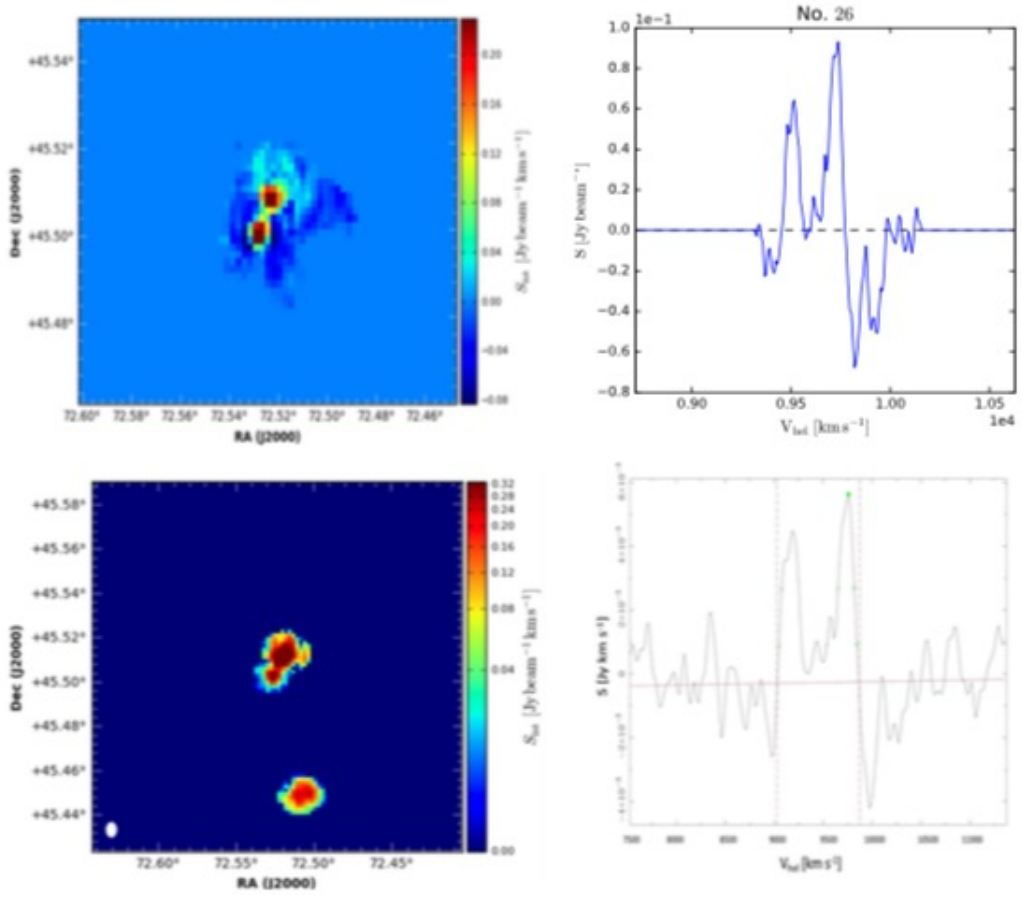


FIGURE 6.10: A SOFIA catalogue with a corresponding crossmatch from the Testbed sample. Top panels: Maps of galaxy no. 26 from SOFIA. Bottom panels: Maps of galaxy no. 3 from Visual.

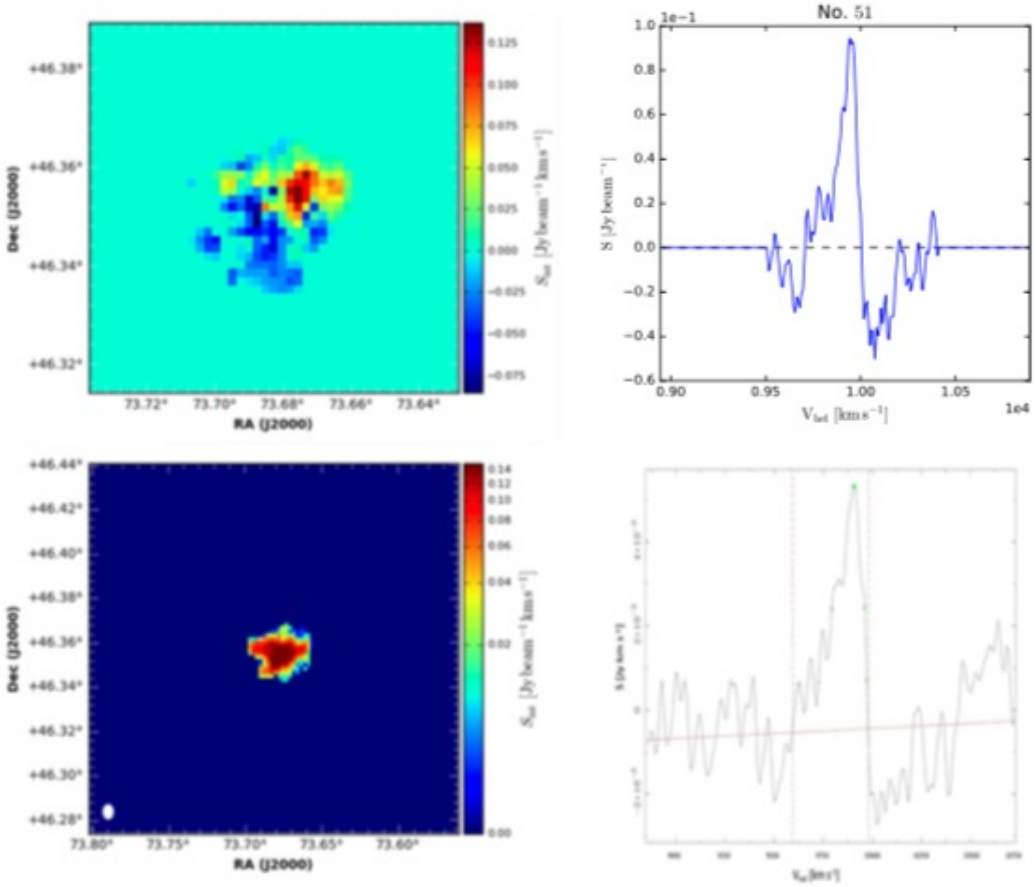


FIGURE 6.11: A SOFIA catalogue with a corresponding crossmatch from the Testbed sample. Top panels: Maps of galaxy no. 51 from SOFIA. Bottom panels: Maps of galaxy no. 4 from Visual.

6.3.2 The relative completeness

The SOFIA catalogue consists of 131 detections, of which 91 are also identified in the Testbed sample. However a significant fraction of 39 detections were not identified by Visual and SA, hence they not in the Testbed sample. There are also 83 Testbed galaxies that are not identified by SOFIA. In this section, I will analyse the detections that are missing in both catalogues and look at their characteristics. Finally will quantify relative reliability and completeness of SOFIA with respect to the Testbed sample.

Figure 6.12 present the histograms of HI parameters. The histograms illustrate where detections with crossmatches lie in comparison to the ones with no crossmatches. The common detections between SOFIA and the Testbed sample ($S \in$ (V-SA)) are shown in red, whereas SOFIA without crossmatches ($S \notin$ (V-SA)) are in grey and blue marks galaxies listed in the Testbed sample but not identified by SOFIA ((V-SA) $\in$ S).

The top left panel present the 50% linewidth histogram. SOFIA has achieved a completeness of $\mathfrak{C} = 61.0\%$ (36/59) and a reliability of $\mathfrak{R} = 87.7\%$ (36/41) at linewidths greater than 150 km s^{-1} but both these quantities drop significantly at lesser linewidths, with $\mathfrak{C} = 47.8\%$ (55/115) and $\mathfrak{R} = 61.1\%$ (55/90).

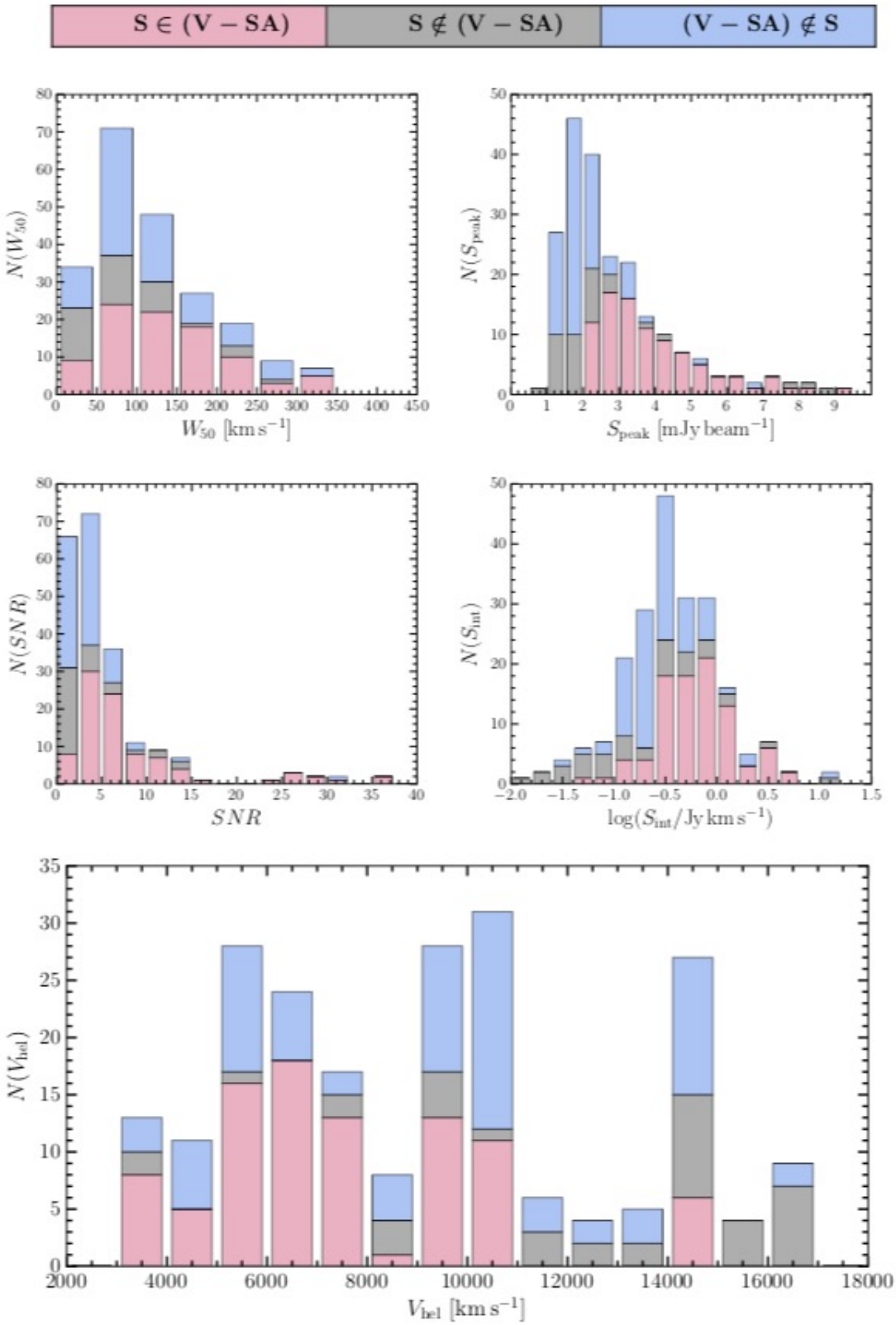


FIGURE 6.12: The distributions of HI parameters of SOFIA detections with and without crossmatches listed in the Testbed sample. S ∈ (V-SA): SOFIA detections with crossmatches. S ∉ (V-SA): SOFIA detection with no crossmatches. (V-SA) ∉ S: Testbed galaxies without crossmatches.

The top right panel shows peak flux histogram. Above the peak flux of $S_{\rm peak} = 2.5$ mJy beam^{-1}, SoFiA reaches a completeness of $\mathfrak{C} = 96.3\%$ (79/82) and reliability of $\mathfrak{R} = 88.8\%$ (79/89). Whereas below this threshold, both the completeness and reliability drop significantly (i.e. $\mathfrak{C} = 13.0\%$ and $\mathfrak{R} = 28.6\%$).

The histogram showing signal-to-noise ratio is presented in the middle left panel. It must be noted that SoFiA performs well for a signal-to-noise ratio greater than 5. Above this threshold it retained a completeness of $\mathfrak{C} = 81.5\%$ (53/65) and $\mathfrak{R} = 84.1\%$ (53/63), compared to $\mathfrak{C} = 34.9\%$ (38/109) and $\mathfrak{R} = 55.9\%$ (38/68) achieved below.

Middle right panel present the integrated flux distribution. For sources with integrated flux greater than $\log\left(S_{\rm int}/{\rm Jy\ km\,s^{-1}}\right) = -0.5$ ($S_{\rm int} = 0.36$ Jy km s^{-1}), SoFiA achieved a completeness of $\mathfrak{C} = 76.9\%$ (70/91) and a reliability of $\mathfrak{R} = 84.3\%$ (70/83). However both these quantities decrease significantly at lesser integrated fluxes, has the returned a completeness of $\mathfrak{C} = 25.3\%$ (21/83) and reliability is $\mathfrak{R} = 43.8\%$ (21/48).

The velocity histogram is shown in the bottom panel. For redshifts in the range of up 2 000 to 10 000 km s^{-1}, SoFiA registered a completeness of $\mathfrak{C} = 63.8\%$ (74/116) and reliability of $\mathfrak{R} = 86.0\%$ (74/86). However at the high end both these quantities drop significantly, with $\mathfrak{C} = 29.8\%$ and $\mathfrak{R} = 37.8\%$.

The distribution plots above show that SOFIA performs well within the redshift of $0.00 - 0.04$, linewidths more than 150 km s^{-1}, SNR greater than 5 and integrated flux of $S_{\rm int} = 0.32$ Jy km s^{-1}. However, further optimisation and fine-tuning of SoFiA parameters are necessary to improve the performance its efficiency at lower HI parameters.

Presented in Fig. 6.13 is the sensitivity plot, the distribution of HI mass as a function of heliocentric velocity, and the histogram of the HI mass. The left panel shows the completeness of SoFiA with respect to the Testbed sample. The galaxies in common between the two catalogues span a mass range of $\log\left(M_{\rm HI}/{\rm M}_\odot\right) = 8.1$ to 9.9 and are centred at $\log\left(M_{\rm HI}/{\rm M}_\odot\right) = 9.4$ with an rms of 0.4. Whereas the SoFiA detections without cross-matches are centred at $\log\left(M_{\rm HI}/{\rm M}_\odot\right) = 8.9$ and have a larger rms of 0.8. The galaxies that SoFiA could not identify have an average of $\log\left(M_{\rm HI}/{\rm M}_\odot\right) = 9.0$ with an rms of 0.4. This shows that SoFiA picks handful of low mass sources which majority are false positives. It must be noted that SoFiA performs fairly good above $\log\left(M_{\rm HI}/{\rm M}_\odot\right) = 8.5$ as it has measured $\mathfrak{C} = 54.0\%$ (87/160) and $\mathfrak{C} = 75.7\%$ (56/74) but there is room for improvement.

SoFiA performs good at R I as it has achieved a completeness of $\mathfrak{C} = 65.1\%$ and reliability of $\mathfrak{R} = 82.0\%$. These are great results considering that a large fraction of the missing galaxies has low SNR ($SNR < 5$) and small linewidths ($W_{50} < 150$ km s^{-1}). Looking at R II, we notice that the completeness drops significantly ($\mathfrak{C} = 15.\%$), however, reliability is still high ($\mathfrak{R} = 60.0\%$). The drop in the completeness is justifiable. If you look at Fig. 5.3, you will notice that majority of the galaxies in R II have linewidths less 100 km s^{-1}. These sources are most likely flagged when applying the reliability filter within SoFiA as they possess similar properties to noise peaks. However, the fact that we managed to retain

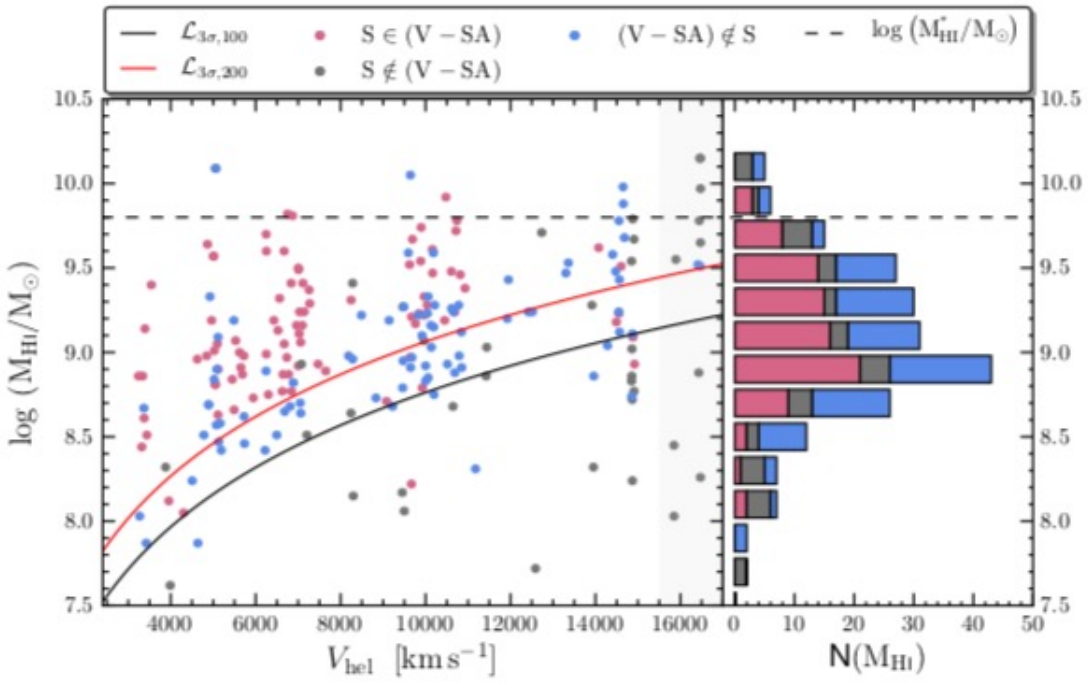


FIGURE 6.13: The distribution of total HI mass as a function heliocentric velocity of SOFIA detections with and without V-SA crossmatches. Left panel: shows the sensitivity plot. Right panel: the histogram of the logarithmic HI mass.

such a high reliability shows the competence and power of SOFIA. At R III the reliability drops down to $\mathfrak{R} = 14.3\%$, while completeness is $\mathfrak{C} = 30.0\%$. SOFIA picked a number of false positive here, however, it was expected as sources in this region are not far removed from the local noise. The table below presents a summary of the efficiency of SOFIA.

TABLE 6.2: Efficiency of SOFIA with respect to the Testbed sample.

	Visual		**SA**	**Reliability**	**Completeness**
	TD	FP	TS	$\mathfrak{R}$ [%]	$\mathfrak{C}$ [%]
RI	82	18	126	82.0	65.1
RII	6	4	38	60.0	15.7
RIII	3	18	10	14.3	30.0
Survey	**91**	**40**	**174**	**69.4**	**52.3**

6.4 Comparing our SOFIA results with literature

A team of Astronomers from the Wallaby survey also conducted a similar performance on SOFIA using real data from ASKAP-12 but with artificial galaxies. They extracted a subcube of 800 × 800 pixels and 800 channels with no notable true HI line emission and populated it with 600 artificial galaxies convolved with the beam of ASKAP. They ran SOFIA on the data cube using the S+C finder with a detection threshold of 3σ and applied the SOFIA reliability filter. Overall, they detected 270 sources of which only 3 were

false positives. Meaning they achieved an impressive 98.9% reliability while measuring a completeness of 44.5%. Whereas in this thesis we achieved a 70% reliability and a 53.9% completeness. It is worth to note that higher completeness measurements may lead to the slight drop in reliability and vice-versa.

Their results show that SoFiA reached 100% completeness at a SNR of 5 and a peak SNR of 2.5 (see Fig. 6 , Westmeier [2018]). Whereas we found a completeness of 81.5% and 84.1% reliability above the SNR of 5. Both these results indicate that SoFiA performs well above 3σ, however, if a user wants to detect sources below this threshold then will have to compromise on reliability.

Chapter 7

Summary

In this project we used an interferometric HI data cube from HIPP-ZoA survey. It encompasses the region where the Perseus-Pisces Supercluster crosses the ZoA. The HIPP-ZoA survey covers a total an area of ~ 9.6 Sq.deg, angular resolution of 23×16 Sq.arcsec, a velocity resolution of $8.24\ \mathrm{km\,s^{-1}}$, an rms noise of $0.40\ \mathrm{mJy\,beam^{-1}}$ and spans a heliocentric velocity range of $2\,400 - 16\,600\ \mathrm{km\,s^{-1}}$.

The primary aim of this project is quantifying the efficiency of a fully automated source finding application, SOFIA, on a spectral data cube with real HI line emission. Unlike in mocked data cubes, the total number of genuine galaxies on the HIPP-ZoA data cube is not known, hence conventional methods of quantifying SOFIA's efficiency are not applicable. However, there is an alternative method that we have opted for in this project. We used two different galaxy identification method (i.e. Visual and Semi-automated) to identify genuine galaxies from the cube. The resulting catalogues were then reconciled and galaxies appearing on both catalogues were put into a Testbed sample. After running SOFIA on the data cube, we then quantified its reliability and completeness with respect to the Testbed sample. Secondly, we looked at dispersions between the HI parameters derived by SOFIA in comparison to the parameters derived by conventional methods that have been used in large HI surveys such as ALFALFA.

Through the Visual method, we identified 194 galaxies. Whereas a Semi-automated method by Ramatsoku et al. [2016] identified a total of 211 galaxies published in the SA catalogue. We then compare Visual catalogue to the SA catalogue. A large fraction of 174 Visual galaxies have crossmatches. This amounts to a relative completeness of $\mathfrak{C} = 82.5\%$ and reliability of $\mathfrak{R} = 89.7\%$. Above the sensitivity curve assuming a detection of 3-sigma, we found a completeness of 86.3% and a reliability 95.5%. We also noticed that above the signal-to-noise ratio of 5, Visual completeness reaches 98.5%. We found that more than 98% of the crossmatches, measured HI parameters that are consistent within the measured errors. All the 174 Visual galaxies with SA crossmatches were then added to the Testbed sample.

We ran SOFIA on a handful of mocked data cubes, with an aim of understanding how it performs on an ideal dataset and which parameters should we focus on when applying

it to the HIPP-ZoA data cube. We then ran SOFIA on the HIPP-ZoA using the following parameters:

- S+C finder, with a detection threshold of three times the noise (3σ).
- Applied a global noise normalisation.
- Applied a reliability filter, where we rejected all detections with reliability less than 0.90.

With the above parameter set-up, we identified a total of 131 galaxy candidates. The detections span peak flux range of 0.80 to 18.70 Jy beam^{-1}, a SNR ranging from 0.14 to 66.97, a 50% linewidth range of $11 - 375$ km s^{-1} and integrated flux range of 5.67 to 9.77 Jy km s^{-1}. Only 3.1% of the detections have HI masses greater than $\log(M^*_{HI})$, the characteristic mass of HI mass function. This is in-line with my expectation, and the steep exponential drop-off towards high Hi mass galaxies above M^*_{HI}.

When comparing the SOFIA catalogue to the Testbed sample, we identified 91 detections with crossmatches. Which is equivalent to a completeness of $\mathfrak{C} = 52.3\%$ and a reliability of $\mathfrak{R} = 69.4\%$. However, if we look at detections above the sensitivity curve assuming a detection of 3-sigma, the completeness rises to $\mathfrak{C} = 65.1\%$ with the reliability reaching $\mathfrak{R} = 82.0\%$. When looking at efficiency as a function of: S_{peak}, W_{50}, SNR and S_{int}, and we made the following findings:

- At peak fluxes above $S_{peak} = 2.5$ mJy km s^{-1}, SOFIA returns a completeness of $\mathfrak{C} = 96.3\%$ and a reliability of $\mathfrak{R} = 88.\%$.
- SOFIA reaches a completeness of $\mathfrak{C} = 61.0\%$ and a reliability of $\mathfrak{R} = 87.7\%$ at linewidths greater than 150 km s^{-1}.
- At SNR greater than 5, we achieved a completeness of $\mathfrak{C} = 81.5\%$ and a reliability of $\mathfrak{R} = 84.1\%$.
- A completeness of $\mathfrak{C} = 76.9\%$ and a reliability of $\mathfrak{R} = 84.3\%$ are achieved at integrated flux greater than $\log(S_{int}/\text{Jy km s}^{-1}) = -0.50$ ($S_{int} = 0.32$ Jy km s^{-1}).

These are great results compared to our initial expectations and are consistent with the results obtained by Westmeier [2018] from the Wallaby survey, where they ran SOFIA on an ASKAP data cube with simulated galaxies. Both the completeness and the reliability are dependent on the noise characteristics of the data. If you have a well-behaved noise as we expect from the SKA pathfinders (e.g. MeerKAT and ASKAP) then SOFIA will perform well.

Lastly, we studied the accuracy of the HI parameters returned by SOFIA such as source centroid, linewidths, signal-to-noise ratio and integrated flux. We made the following findings:

1. The returned source centroids show a high degree of accuracy within statistical errors.

2. The heliocentric velocities show acceptable dispersions of less than 100 $\mathrm{km\,s^{-1}}$.

3. For a large fraction of 94.5% sources, SoFiA measurements in parameters such as linewidths, integrated flux and HI mass show consistency within deviations less than $3\times$rms.

A few sources show larger deviations in returned parameters. A closer look at these sources revealed that the greater deviations are the results of three phenomenon: 1) satellites being merged with the parent source, 2) sources broken into multiple detections. The first phenomenon is common in regions of high density as you find a number of HI bridges due to strong tidal interactions. The second phenomenon result from faint edge-on galaxies characterised by double-horn global profiles which are largely buried in the noise. Hence the two horns are detected as multiple sources.

This work provides a fairly good understanding of how SoFiA performs on real data and what level of efficiency is to be expected on forthcoming large HI surveys. It must be noted that SoFiA is highly efficient above 3σ, however, more tests and parameter fine-tuning is necessary to detect the genuine sources below 3σ while preserving the higher reliability. Hence in the near future, I would like to: (i) broaden the size of fine-tuned parameters, (ii) test other source finder and filters within SoFiA, (iii) run SoFiA on a different dataset (e.g. MeerKAT data). Then produce a manual advising researchers on which combination of parameters are recommended for certain types of datasets.

Bibliography

Andrew J Baker, Sarah Blyth, Benne W Holwerda, et al. Laduma: Looking at the distant universe with the meerkat array. In *American Astronomical Society Meeting Abstracts*, volume 231, 2018.

D. G. Barnes, L. Staveley-Smith, W. J. G. de Blok, T. Oosterloo, I. M. Stewart, A. E. Wright, G. D. Banks, R. Bhathal, P. J. Boyce, M. R. Calabretta, M. J. Disney, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, P. Te Lintel Hekkert, P. A. Henning, H. Jerjen, S. Juraszek, M. J. Kesteven, V. A. Kilborn, P. M. Knezek, B. Koribalski, R. C. Kraan-Korteweg, D. F. Malin, M. Marquarding, R. F. Minchin, J. R. Mould, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, A. Schröder, F. Stootman, R. L. Webster, W. E. Wilson, and T. Ye. The h i parkes all sky survey: southern observations, calibration and robust imaging. *Monthly Notices of the Royal Astronomical Society*, 322(3):486–498, 2001. doi: 10.1046/j.1365-8711.2001.04102.x. URL http://dx.doi.org/10.1046/j.1365-8711.2001.04102.x.

RS Booth, WJG De Blok, JL Jonas, and B Fanaroff. Meerkat key project science, specifications, and proposals. *arXiv preprint arXiv:0910.2935*, 2009.

Bernard F Burke and Francis Graham-Smith. *An introduction to radio astronomy*. Cambridge University Press, 2009.

Christopher Carilli and Steve Rawlings. *Msc dissertation*, 2003.

Christopher Carilli and Steve Rawlings. Science with the square kilometer array: motivation, key science projects, standards and assumptions. *arXiv preprint astro-ph/0409274*, 2004.

ME Cluver, TH Jarrett, RC Kraan-Korteweg, BS Koribalski, PN Appleton, J Melbourne, B Emonts, and PA Woudt. Active disk building in a local h i-massive lirg: the synergy between gas, dust, and star formation. *The Astrophysical Journal*, 725(2):1550, 2010.

Luca Cortese, Robert Frank Minchin, Robbie Richard Auld, Jonathan Ivor Davies, B Catinella, E Momjian, JL Rosenberg, Rhys Taylor, G Gavazzi, K O'Neil, et al. The arecibo galaxy environment survey–ii. ah i view of the abell cluster 1367 and its outskirts. *Monthly Notices of the Royal Astronomical Society*, 383(4):1519–1537, 2008.

WJG de Blok, EAK Adams, P Amram, E Athanassoula, I Bagetakos, C Balkowski, MA Bershady, R Beswick, F Bigiel, S-L Blyth, et al. An overview of the mhongoose survey: Observing nearby galaxies with meerkat. *arXiv preprint arXiv:1709.08458*, 2017.

Jennifer L Donley, L Staveley-Smith, RC Kraan-Korteweg, JM Islas-Islas, A Schröder, PA Henning, B Koribalski, S Mader, and I Stewart. The hi parkes zone of avoidance survey: The northern extension. *The Astronomical Journal*, 129(1):220, 2005.

Alan R Duffy, Martin J Meyer, Lister Staveley-Smith, Maksym Bernyk, Darren J Croton, Bärbel S Koribalski, Derek Gerstmann, and Stefan Westerlund. Predictions for askap neutral hydrogen surveys. *Monthly Notices of the Royal Astronomical Society*, 426(4):3385–3402, 2012.

Harold Irving Ewen and Edward Mills Purcell. Observation of a line in the galactic radio spectrum: Radiation from galactic hydrogen at 1,420 mc./sec. *Nature*, 168(4270):356, 1951.

Lars Flöer and Benjamin Winkel. 2d–1d wavelet reconstruction as a tool for source finding in spectroscopic imaging surveys. *Publications of the Astronomical Society of Australia*, 29 (3):244–250, 2012.

Lars Flöer, Benjamin Winkel, and Jürgen Kerp. Source finding, parametrization, and classification for the extragalactic effelsberg-bonn h i survey. *Astronomy & Astrophysics*, 569: A101, 2014.

P Focardi, B Marano, and G Vettolani. The north extension of the perseus supercluster. *Astronomy and Astrophysics*, 136:178–180, 1984.

Riccardo Giovanelli and Martha P Haynes. Extragalactic hi surveys. *The Astronomy and Astrophysics Review*, 24(1):1, 2016.

Riccardo Giovanelli, Martha P Haynes, Brian R Kent, Philip Perillat, Amelie Saintonge, Noah Brosch, Barbara Catinella, G Lyle Hoffman, Sabrina Stierwalt, Kristine Spekkens, et al. The arecibo legacy fast alfa survey. i. science goals, survey design, and strategy. *The astronomical journal*, 130(6):2598, 2005.

Richard Gooch. Karma users manual. *ATNF, Sydney*, 2006.

Martha P. Haynes, Riccardo Giovanelli, Ann M. Martin, Kelley M. Hess, Amélie Saintonge, Elizabeth A. K. Adams, Gregory Hallenbeck, G. Lyle Hoffman, Shan Huang, Brian R. Kent, Rebecca A. Koopmann, Emmanouil Papastergis, Sabrina Stierwalt, Thomas J. Balonek, David W. Craig, Sarah J. U. Higdon, David A. Kornreich, Jeffrey R. Miller, Aileen A. O'Donoghue, Ronald P. Olowin, Jessica L. Rosenberg, Kristine Spekkens, Parker Troischt, and Eric M. Wilcots. The arecibo legacy fast alfa survey: The α.40 h i source catalog, its characteristics and their impact on the derivation of the h i mass function. *The Astronomical Journal*, 142(5):170, 2011. URL http://stacks.iop.org/1538-3881/142/i=5/a=170.

PA Henning, CM Springob, F Day, R Minchin, E Momjian, B Catinella, E Muller, B Koribalski, K Masters, C Pantoja, et al. The alfa zone of avoidance survey. In *AIP Conference Proceedings*, volume 1035, pages 246–248. AIP, 2008.

K. G. Begeman W. Zwitser J. M. Van der Hulst, J. P. Terlouw and P. R. Roelfsema. Astronomical data analysis software and systems i. 25, 1992.

S. Johnston, R. Taylor, M. Bailes, N. Bartel, C. Baugh, M. Bietenholz, C. Blake, R. Braun, J. Brown, S. Chatterjee, J. Darling, A. Deller, R. Dodson, P. Edwards, R. Ekers, S. Ellingsen, I. Feain, B. Gaensler, M. Haverkorn, G. Hobbs, A. Hopkins, C. Jackson, C. James, G. Joncas, V. Kaspi, V. Kilborn, B. Koribalski, R. Kothes, T. Landecker, A. Lenc, J. Lovell, J.-P. Macquart, R. Manchester, D. Matthews, N. McClure-Griffiths, R. Norris, U.-L. Pen, C. Phillips, C. Power, R. Protheroe, E. Sadler, B. Schmidt, I. Stairs, L. Staveley-Smith, J. Stil, S. Tingay, A. Tzioumis, M. Walker, J. Wall, and M. Wolleben. Science with askap. *Experimental Astronomy*, 22(3):151–273, Dec 2008. ISSN 1572-9508. doi: 10.1007/s10686-008-9124-7. URL https://doi.org/10.1007/s10686-008-9124-7.

Russell Jurek. The characterised noise h i source finder: Detecting h i galaxies using a novel implementation of matched filtering. *Publications of the Astronomical Society of Australia*, 29(3):251–261, 2012.

V. A. Kilborn. Phd thesis. *A&A*, 2001.

B. S. Koribalski, L. Staveley-Smith, V. A. Kilborn, S. D. Ryder, R. C. Kraan-Korteweg, E. V. Ryan-Weber, R. D. Ekers, H. Jerjen, P. A. Henning, M. E. Putman, M. A. Zwaan, W. J. G. de Blok, M. R. Calabretta, M. J. Disney, R. F. Minchin, R. Bhathal, P. J. Boyce, M. J. Drinkwater, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, S. Juraszek, M. J. Kesteven, P. M. Knezek, S. Mader, M. Marquarding, M. Meyer, J. R. Mould, T. Oosterloo, J. O'Brien, R. M. Price, E. M. Sadler, A. Schröder, I. M. Stewart, F. Stootman, M. Waugh, B. E. Warren, R. L. Webster, and A. E. Wright. The 1000 brightest hipass galaxies: H i properties. *The Astronomical Journal*, 128(1):16, 2004. URL http://stacks.iop.org/1538-3881/128/i=1/a=16.

K Kovač, TA Oosterloo, and JM Van Der Hulst. A blind h i survey in the canes venatici region. *Monthly Notices of the Royal Astronomical Society*, 400(2):743–765, 2009.

RC Kraan-Korteweg, AJ Loan, WB Burton, O Lahav, HC Ferguson, PA Henning, and D Lynden-Bell. Discovery of a nearby spiral galaxy behind the milky way. *Nature*, 372(6501):77, 1994.

Henric Krawczynski. Spectroscopic deprojection analysis of chandra data of the galaxy cluster 3c 129. *The Astrophysical Journal Letters*, 569(1):L27, 2002.

RK Lutz. An algorithm for the real time analysis of digitised images. *The computer journal*, 23(3):262–269, 1980.

Travis P McIntyre, Patricia A Henning, Robert F Minchin, Emmanuel Momjian, and Zhon Butcher. The alfa zoa deep survey: First results. *The Astronomical Journal*, 150(1):28, 2015.

Martin Meyer. Exploring the hi universe with askap. *arXiv preprint arXiv:0912.2167*, 2009.

Martin J Meyer, Martin A Zwaan, Rachel L Webster, Lister Staveley-Smith, E Ryan-Weber, Michael John Drinkwater, David G Barnes, Matt Howlett, Virginia A Kilborn, J Stevens, et al. The hipass catalogue–i. data presentation. *Monthly Notices of the Royal Astronomical Society*, 350(4):1195–1209, 2004.

Attila Popping, Russell Jurek, Tobias Westmeier, Paolo Serra, L Flöer, Martin Meyer, and Baerbel Koribalski. Comparison of potential askap h i survey source finders. *Publications of the Astronomical Society of Australia*, 29(3):318–339, 2012.

M Ramatsoku, MAW Verheijen, RC Kraan-Korteweg, GIG Józsa, AC Schröder, TH Jarrett, EC Elson, W van Driel, WJG de Blok, and PA Henning. The wsrt zoa perseus-pisces filament wide-field h i imaging survey–i. hi catalogue and atlas. *Monthly Notices of the Royal Astronomical Society*, 460(1):923–941, 2016.

Jessica L Rosenberg and Stephen E Schneider. The arecibo dual-beam survey: Arecibo* and vla** observations. *The Astrophysical Journal Supplement Series*, 130(1):177, 2000.

J Braatz S Stierwalt, A Peck and A Bemis. Introduction to radio interferometry. 2016. URL https://science.nrao.edu/facilities/alma/naasc-workshops/nrao-cd-wm16/BasicsInterf_Madison.pdf.

Amélie Saintonge. The arecibo legacy fast alfa survey. iv. strategies for signal identification and survey catalog reliability. *The Astronomical Journal*, 133(5):2087, 2007.

R Sault and N Killeen. Multichannel image reconstruction, image analysis and display (miriad) users guide, atnf, 1996.

P Serra, R Jurek, and L Flöer. Using negative detections to estimate source-finder reliability. *Publications of the Astronomical Society of Australia*, 29(3):296–300, 2012a.

Paolo Serra, Tom Oosterloo, Raffaella Morganti, Katherine Alatalo, Leo Blitz, Maxime Bois, Frédéric Bournaud, Martin Bureau, Michele Cappellari, Alison F Crocker, et al. The atlas3d project–xiii. mass and morphology of h i in early-type galaxies as a function of environment. *Monthly Notices of the Royal Astronomical Society*, 422(3):1835–1862, 2012b.

Paolo Serra, Tobias Westmeier, Nadine Giese, Russell Jurek, Lars Flöer, Attila Popping, Benjamin Winkel, Thijs van der Hulst, Martin Meyer, Bärbel S Koribalski, et al. Sofia: a flexible source finder for 3d spectral line data. *Monthly Notices of the Royal Astronomical Society*, 448(2):1922–1929, 2015.

Paolo Serra, WJG de Blok, GL Bryan, S Colafrancesco, R-J Dettmar, BS Frank, F Govoni, GIG Józsa, RC Kraan-Korteweg, SI Loubser, et al. The meerkat fornax survey. *arXiv preprint arXiv:1709.01289*, 2017.

J-L Starck, Jalal M Fadili, Seth Digel, Bo Zhang, and Jim Chiang. Source detection using a 3d sparse representation: application to the fermi gamma-ray space telescope. *Astronomy & Astrophysics*, 504(2):641–652, 2009.

L Staveley-Smith, RC Kraan-Korteweg, AC Schröder, PA Henning, BS Koribalski, IM Stewart, and G Heald. The parkes hi zone of avoidance survey. *The Astronomical Journal*, 151(3):52, 2016.

Adoalbjörn Stefánsson, N Končar, and Antonia J Jones. A note on the gamma test. *Neural Computing & Applications*, 5(3):131–133, 1997.

M Storey. Introduction to radio interferometry. 2002.

A. R. Taylor. The square kilometre array. *Proceedings of the International Astronomical Union*, 8(S291):337–341, 2012. doi: 10.1017/S1743921312024039.

J. M. van der Hulst, T. S. van Albada, and R. Sancisi. The Westerbork HI Survey of Irregular and Spiral Galaxies, WHISP. In J. E. Hibbard, M. Rupen, and J. H. van Gorkom, editors, *Gas and Galaxy Evolution*, volume 240 of *Astronomical Society of the Pacific Conference Series*, page 451, 2001.

M. A. W. Verheijen, N. Trentham, R. B. Tully, and M. A. Zwaan. The H I Mass Function in the Ursa Major Cluster. In R. C. Kraan-Korteweg, P. A. Henning, and H. Andernach, editors, *Mapping the Hidden Universe: The Universe behind the Mily Way - The Universe in HI*, volume 218 of *Astronomical Society of the Pacific Conference Series*, page 263, 2000.

M. A. W. Verheijen, T. A. Oosterloo, W. A. van Cappellen, L. Bakker, M. V. Ivashina, and J. M. van der Hulst. Apertif, a focal plane array for the wsrt. In R Minchin and E Momjian, editors, *EVOLUTION OF GALAXIES THROUGH THE NEUTRAL HYDROGEN WINDOW*, AIP CONFERENCE PROCEEDINGS, pages 265–271. AMER INST PHYSICS, 2008. ISBN 978-0-7354-0558-5.

T. Westmeier. Testing sofia on askap early science data. 9, 2018. URL https://wallaby-survey.org/wp-content/uploads/2018/04/WallabyNewsletter9.pdf.

T. Westmeier, A. Popping, and P. Serra. Basic testing of the duchamp source finder. *Publications of the Astronomical Society of Australia*, 29(3):276–295, 2012. doi: 10.1071/AS11041.

Matthew T. Whiting. duchamp: a 3d source finder for spectral-line data. *Monthly Notices of the Royal Astronomical Society*, 421(4):3242–3256, 2012. doi: 10.1111/j.1365-2966.2012.20548.x. URL http://dx.doi.org/10.1111/j.1365-2966.2012.20548.x.

M. S. Yun, P. T. P. Ho, and K. Y. Lo. A high-resolution image of atomic hydrogen in the m81 group of galaxies. *Nature*, 372:530–532, 1994. URL https://www.nature.com/articles/372530a0.

M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster. The hipass catalogue: ω h i and environmental effects on the h i mass function of galaxies. *Monthly Notices of the Royal Astronomical Society: Letters*, 359(1):L30–L34, 2005. doi: 10.1111/j.1745-3933.2005.00029.x. URL http://dx.doi.org/10.1111/j.1745-3933.2005.00029.x.

Martin A. Zwaan, Frank H. Briggs, David Sprayberry, and Ertu Sorar. The h i mass function of galaxies from a deep survey in the 21 centimeter line. *The Astrophysical Journal*, 490(1):173, 1997. URL http://stacks.iop.org/0004-637X/490/i=1/a=173.

Appendix A

Visual catalogue

TABLE A.1: Visual method galaxy catalogue consisting of 194 detections.

No.	Source ID	Subcube	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	log (M_{HI})	SAC
			deg	deg	km s^{-1}	km s^{-1}	km s^{-1}	mJy beam^{-1}	Jy km s^{-1}	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	J044238.11 + 442618.1	9	160.11	−1.11	10696 ± 57	203 ± 71	171 ± 228	1.11 ± 0.40	0.04 ± 0.26	0.27	153	8.29	–
2	J044313.85 + 452731.7	3	159.41	−0.36	5480 ± 03	276 ± 09	245 ± 12	1.11 ± 0.40	0.60 ± 0.08	3.92	78	8.94	–
3	J044426.83 + 455117.4	3	159.25	+0.06	5487 ± 01	279 ± 03	265 ± 04	1.61 ± 0.41	1.07 ± 0.07	6.68	78	9.19	Y
4	J044522.55 + 454436.5	6	159.44	+0.11	5113 ± 02	298 ± 06	281 ± 08	2.88 ± 0.43	0.90 ± 0.11	5.42	73	9.05	Y
5	J044523.78 + 451929.3	6	159.76	−0.16	6811 ± 15	142 ± 45	63 ± 60	2.00 ± 0.41	0.21 ± 0.09	2.74	97	8.68	Y
6	J044540.46 + 454057.9	6	159.52	+0.11	9647 ± 07	145 ± 21	131 ± 28	2.04 ± 0.41	0.18 ± 0.14	1.62	138	8.91	Y
7	J044542.33 + 455043.6	3	159.40	+0.22	14564 ± 13	171 ± 39	71 ± 52	2.18 ± 0.41	0.27 ± 0.09	3.19	208	9.43	Y
8	J044544.03 + 442055.1	9	160.54	−0.75	10701 ± 04	198 ± 12	178 ± 16	2.37 ± 0.42	0.39 ± 0.12	2.98	153	9.33	Y
9	J044546.66 + 444858.0	6	160.19	−0.44	5128 ± 09	92 ± 27	55 ± 36	1.96 ± 0.41	0.23 ± 0.08	3.17	73	8.47	Y
10	J044550.65 + 443608.8	6	160.36	−0.57	16466 ± 05	88 ± 15	73 ± 20	1.12 ± 0.40	0.25 ± 0.07	2.93	235	9.51	Y
11	J044557.41 + 450758.1	6	159.97	−0.21	6785 ± 03	84 ± 09	70 ± 12	2.78 ± 0.42	0.48 ± 0.07	5.81	97	9.03	Y
12	J044559.21 + 450108.1	6	160.06	−0.28	12513 ± 12	148 ± 36	86 ± 48	1.45 ± 0.41	0.23 ± 0.08	2.51	179	9.24	Y
13	J044601.97 + 443442.8	6	160.40	−0.56	6265 ± 01	117 ± 03	99 ± 04	4.43 ± 0.46	2.42 ± 0.05	24.64	90	9.66	Y
14	J044628.17 + 451719.8	6	159.91	−0.04	10199 ± 16	137 ± 48	109 ± 64	1.75 ± 0.41	0.11 ± 0.15	1.10	146	8.75	Y
15	J044632.07 + 444313.6	6	160.35	−0.40	9846 ± 03	148 ± 09	131 ± 12	2.72 ± 0.42	0.42 ± 0.10	3.70	141	9.29	–
16	J044632.69 + 452156.9	6	159.86	+0.02	9722 ± 03	226 ± 09	209 ± 12	2.40 ± 0.42	0.53 ± 0.11	3.73	139	9.38	Y
17	J044635.70 + 441920.5	9	160.66	−0.65	9978 ± 07	132 ± 21	101 ± 28	2.37 ± 0.42	0.28 ± 0.10	2.79	143	9.12	Y
18	J044638.36 + 444632.6	6	160.32	−0.35	9893 ± 08	186 ± 24	45 ± 32	1.65 ± 0.41	0.54 ± 0.04	8.12	141	9.40	–
19	J044643.36 + 442030.5	9	160.66	−0.62	5637 ± 03	189 ± 09	171 ± 12	3.16 ± 0.43	0.55 ± 0.11	4.28	81	8.93	Y
20	J044646.08 + 444742.6	6	160.32	−0.32	5694 ± 03	161 ± 09	145 ± 12	2.85 ± 0.42	0.45 ± 0.11	3.83	81	8.85	Y
21	J044652.58 + 440918.6	9	160.82	−0.72	7488 ± 07	163 ± 21	138 ± 28	2.46 ± 0.42	0.26 ± 0.13	2.20	107	8.84	Y
22	J044655.23 + 444643.7	6	160.35	−0.31	16411 ± 11	121 ± 33	53 ± 44	1.19 ± 0.40	0.26 ± 0.06	3.58	234	9.52	Y
23	J044655.73 + 450851.3	6	160.07	−0.07	10790 ± 15	88 ± 45	36 ± 60	1.66 ± 0.41	0.17 ± 0.07	2.86	154	8.98	Y
24	J044656.12 + 450941.0	6	160.06	−0.06	10780 ± 08	150 ± 24	89 ± 32	2.15 ± 0.41	0.34 ± 0.08	3.69	154	9.28	Y
25	J044702.03 + 442455.3	9	160.64	−0.53	10735 ± 01	336 ± 03	318 ± 04	3.44 ± 0.44	1.37 ± 0.10	7.81	153	9.88	Y
26	J044705.19 + 450311.1	6	160.16	−0.11	10124 ± 08	169 ± 24	122 ± 32	1.39 ± 0.41	0.29 ± 0.09	2.68	145	9.16	Y
27	J044707.86 + 453504.3	6	159.76	+0.24	4976 ± 01	216 ± 03	201 ± 04	4.99 ± 0.47	1.13 ± 0.11	8.11	71	9.13	Y
28	J044717.47 + 441702.0	9	160.77	−0.58	5087 ± 04	374 ± 12	316 ± 16	1.34 ± 0.41	0.63 ± 0.09	3.61	73	8.90	Y
29	J044721.53 + 435450.3	9	161.06	−0.81	10057 ± 04	199 ± 12	185 ± 16	2.03 ± 0.41	0.32 ± 0.12	2.36	144	9.19	Y
30	J044724.27 + 443656.5	6	160.53	−0.35	6222 ± 18	125 ± 54	74 ± 72	1.58 ± 0.41	0.14 ± 0.10	1.65	89	8.42	Y
31	J044726.80 + 444732.8	6	160.40	−0.23	7279 ± 09	110 ± 27	63 ± 36	2.89 ± 0.43	0.28 ± 0.09	3.54	104	8.85	Y
32	J044727.37 + 445355.4	6	160.32	−0.16	5505 ± 06	70 ± 18	48 ± 24	3.22 ± 0.43	0.28 ± 0.09	4.03	79	8.60	Y

TABLE A.1: Visual method galaxy catalogue consisting of 194 detections.

No.	Source ID	Subcube	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	$\log(M_{HI})$	SAC
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
33	J044730.34 + 434819.3	9	161.16	−0.86	11177 ± 37	78 ± 11	65 ± 148	1.52 ± 0.41	0.03 ± 0.19	0.43	160	8.31	Y
34	J044733.99 + 455923.3	3	159.50	+0.56	9916 ± 07	121 ± 21	94 ± 28	2.26 ± 0.42	0.27 ± 0.10	2.77	142	9.10	Y
35	J044735.48 + 451622.1	6	160.05	+0.10	3268 ± 08	154 ± 24	129 ± 32	1.17 ± 0.40	0.21 ± 0.10	1.88	47	8.03	Y
36	J044742.67 + 442003.0	9	160.78	−0.49	4859 ± 03	143 ± 09	114 ± 12	4.04 ± 0.45	0.69 ± 0.09	6.58	69	8.90	Y
37	J044743.94 + 460446.7	3	159.45	+0.64	12422 ± 01	132 ± 03	132 ± 04	1.72 ± 0.41	0.23 ± 0.11	2.07	177	9.24	Y
38	J044751.67 + 435838.7	9	161.07	−0.70	8188 ± 06	92 ± 18	66 ± 24	1.64 ± 0.41	0.30 ± 0.07	3.71	117	8.98	Y
39	J044753.97 + 443903.4	6	160.56	−0.26	10061 ± 13	106 ± 39	77 ± 52	1.77 ± 0.41	0.15 ± 0.11	1.69	144	8.85	Y
40	J044809.23 + 461619.8	3	159.35	+0.82	4021 ± 24	69 ± 72	37 ± 96	1.72 ± 0.41	0.08 ± 0.10	1.38	57	7.81	–
41	J044809.61 + 443641.5	6	160.62	−0.25	9858 ± 05	115 ± 15	99 ± 20	2.60 ± 0.42	0.27 ± 0.11	2.77	141	9.10	Y
42	J044810.42 + 450330.3	6	160.28	+0.04	9833 ± 06	264 ± 18	232 ± 24	2.13 ± 0.41	0.35 ± 0.13	2.35	140	9.22	Y
43	J044819.32 + 440203.3	9	161.08	−0.60	4781 ± 11	163 ± 33	81 ± 44	2.15 ± 0.41	0.29 ± 0.09	3.32	68	8.51	Y
44	J044838.05 + 454514.7	6	159.80	+0.55	5731 ± 04	140 ± 12	109 ± 16	2.94 ± 0.43	0.56 ± 0.09	5.40	82	8.94	Y
45	J044839.91 + 452421.1	6	160.07	+0.33	10941 ± 04	171 ± 12	153 ± 16	2.96 ± 0.43	0.35 ± 0.13	2.86	156	9.30	Y
46	J044845.09 + 435632.8	9	161.20	−0.60	3392 ± 02	37 ± 06	24 ± 08	3.58 ± 0.44	0.59 ± 0.04	12.26	48	8.52	Y
47	J044852.05 + 445005.8	6	160.53	−0.01	14284 ± 23	139 ± 69	86 ± 92	1.22 ± 0.40	0.11 ± 0.11	1.21	204	9.04	Y
48	J044854.70 + 454915.1	3	159.78	+0.63	10103 ± 13	89 ± 39	72 ± 52	1.51 ± 0.41	0.12 ± 0.11	1.38	144	8.75	–
49	J044858.52 + 451259.7	6	160.25	+0.25	10842 ± 11	263 ± 33	208 ± 44	1.24 ± 0.40	0.23 ± 0.12	1.63	155	9.12	Y
50	J044858.72 + 441945.4	9	160.93	−0.32	14600 ± 14	190 ± 42	157 ± 56	1.64 ± 0.41	0.14 ± 0.15	1.14	209	9.16	Y
51	J044908.85 + 443040.3	9	160.81	−0.18	10657 ± 09	204 ± 27	135 ± 36	1.88 ± 0.41	0.32 ± 0.10	2.75	152	9.24	Y
52	J044912.56 + 441400.2	9	161.03	−0.35	14521 ± 05	349 ± 15	266 ± 20	1.24 ± 0.40	0.67 ± 0.08	4.18	207	9.83	–
53	J044913.24 + 442503.0	9	160.89	−0.23	8290 ± 03	289 ± 09	237 ± 12	2.32 ± 0.42	0.73 ± 0.10	4.83	118	9.38	Y
54	J044931.16 + 435848.0	9	161.26	−0.47	9224 ± 01	174 ± 03	187 ± 04	1.45 ± 0.41	0.12 ± 0.17	0.88	132	8.68	Y
55	J044936.05 + 451042.4	6	160.35	+0.31	6635 ± 09	81 ± 27	41 ± 36	3.36 ± 0.43	0.24 ± 0.09	3.80	95	8.71	Y
56	J044936.65 + 445248.5	6	160.58	+0.12	3970 ± 12	65 ± 36	36 ± 48	2.58 ± 0.42	0.16 ± 0.09	2.64	57	8.07	Y
57	J044943.11 + 460233.7	3	159.70	+0.88	14918 ± 03	78 ± 09	57 ± 12	17.88 ± 0.56	0.52 ± 0.10	6.95	213	9.74	Y
58	J044958.16 + 460310.7	3	159.72	+0.92	4924 ± 01	188 ± 03	174 ± 04	6.50 ± 0.52	1.83 ± 0.09	14.11	70	9.33	Y
59	J045001.31 + 452655.3	6	160.19	+0.54	9594 ± 01	123 ± 03	110 ± 04	3.26 ± 0.43	0.88 ± 0.07	8.48	137	9.59	Y
60	J045001.38 + 444925.9	6	160.67	+0.14	3433 ± 06	102 ± 18	37 ± 24	12.70 ± 0.42	0.49 ± 0.05	8.21	49	8.45	Y
61	J045005.69 + 453041.1	6	160.15	+0.59	9643 ± 02	292 ± 06	69 ± 08	3.33 ± 0.43	2.50 ± 0.03	30.52	138	10.05	Y
62	J045006.14 + 450324.8	6	160.50	+0.30	14884 ± 11	68 ± 33	32 ± 44	3.05 ± 0.43	0.19 ± 0.08	3.43	213	9.31	Y
63	J045011.74 + 440847.6	9	161.21	−0.27	5532 ± 03	137 ± 09	106 ± 12	4.66 ± 0.46	0.71 ± 0.09	7.04	79	9.02	Y
64	J045023.36 + 442214.2	9	161.06	−0.10	5726 ± 09	121 ± 27	77 ± 36	2.11 ± 0.41	0.27 ± 0.09	3.08	82	8.62	Y

TABLE A.1: Visual method galaxy catalogue consisting of 194 detections.

No.	Source ID	Subcube	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	$\log(M_{HI})$	SAC
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
65	J045024.88 + 451137.0	6	160.43	+0.43	4629 ± 04	110 ± 12	69 ± 16	4.25 ± 0.45	0.56 ± 0.08	6.89	66	8.76	Y
66	J045033.08 + 435929.7	9	161.37	−0.32	5736 ± 18	167 ± 54	79 ± 72	1.60 ± 0.41	0.18 ± 0.09	2.10	82	8.46	Y
67	J045035.40 + 442227.9	9	161.08	−0.07	4320 ± 11	52 ± 33	23 ± 44	3.31 ± 0.43	0.17 ± 0.07	3.62	62	8.19	Y
68	J045044.17 + 443921.8	6	160.88	+0.13	14551 ± 08	149 ± 24	132 ± 32	1.28 ± 0.41	0.17 ± 0.11	1.52	208	9.24	Y
69	J045103.54 + 443512.4	5	160.97	+0.13	4994 ± 03	87 ± 09	36 ± 12	26.09 ± 0.50	0.92 ± 0.06	15.53	71	9.04	Y
70	J045103.99 + 443603.1	5	160.96	+0.14	5028 ± 01	190 ± 03	169 ± 04	6.78 ± 0.52	3.18 ± 0.07	24.79	72	9.59	Y
71	J045128.48 + 462002.9	2	159.67	+1.30	5076 ± 09	140 ± 27	76 ± 36	2.07 ± 0.41	0.30 ± 0.08	3.50	73	8.57	Y
72	J045131.44 + 442911.8	8	161.10	+0.13	5114 ± 02	311 ± 06	289 ± 08	3.12 ± 0.43	0.97 ± 0.11	5.77	73	9.09	Y
73	J045135.58 + 461930.3	2	159.69	+1.31	7012 ± 04	95 ± 12	79 ± 16	2.37 ± 0.42	0.32 ± 0.09	3.68	100	8.88	Y
74	J045137.17 + 452430.2	5	160.40	+0.73	9642 ± 12	67 ± 36	31 ± 48	3.47 ± 0.44	0.17 ± 0.09	3.17	138	8.89	Y
75	J045145.00 + 443638.3	5	161.03	+0.24	5065 ± 01	269 ± 03	241 ± 04	5.37 ± 00.48	10.02 ± 0.04	65.47	72	10.09	Y
76	J045157.06 + 450317.0	5	160.71	+0.55	4506 ± 23	184 ± 69	50 ± 92	1.66 ± 0.41	0.18 ± 0.08	2.57	64	8.24	Y
77	J045200.04 + 435345.4	8	161.61	−0.18	14870 ± 13	168 ± 39	148 ± 52	1.94 ± 0.41	0.12 ± 0.17	1.01	212	9.11	Y
78	J045202.61 + 435408.3	8	161.61	−0.17	13959 ± 15	45 ± 45	34 ± 60	3.53 ± 0.44	0.08 ± 0.14	1.36	199	8.86	Y
79	J045202.61 + 435408.3	8	161.61	−0.17	11923 ± 07	157 ± 21	136 ± 28	1.82 ± 0.41	0.23 ± 0.12	2.02	170	9.20	Y
80	J045207.14 + 434922.2	8	161.68	−0.21	7224 ± 01	334 ± 03	315 ± 04	3.43 ± 0.44	1.55 ± 0.09	8.86	103	9.59	Y
81	J045216.83 + 450829.2	5	160.68	+0.65	6665 ± 05	154 ± 15	137 ± 20	3.06 ± 0.43	0.29 ± 0.14	2.55	95	8.80	Y
82	J045220.74 + 454140.1	5	160.26	+1.01	6440 ± 02	192 ± 06	176 ± 08	4.08 ± 0.45	0.73 ± 0.11	5.61	92	9.17	Y
83	J045222.83 + 452759.5	5	160.44	+0.87	5964 ± 03	103 ± 09	86 ± 12	3.27 ± 0.43	0.48 ± 0.09	5.21	85	8.91	Y
84	J045225.94 + 444544.2	5	160.99	+0.43	5130 ± 11	98 ± 33	62 ± 44	2.41 ± 0.42	0.19 ± 0.10	2.40	73	8.37	Y
85	J045230.89 + 461236.7	2	159.88	+1.36	7090 ± 04	143 ± 12	119 ± 16	3.47 ± 0.44	0.44 ± 0.11	4.07	101	9.03	Y
86	J045254.80 + 460249.4	2	160.05	+1.31	7670 ± 07	100 ± 21	86 ± 28	2.34 ± 0.42	0.18 ± 0.12	1.93	110	8.70	Y
87	J045325.17 + 440618.7	8	161.61	+0.15	7084 ± 08	72 ± 24	46 ± 32	2.17 ± 0.41	0.23 ± 0.07	3.52	101	8.75	–
88	J045329.63 + 443929.6	5	161.19	+0.51	10135 ± 14	113 ± 42	41 ± 56	1.79 ± 0.41	0.22 ± 0.07	3.44	145	9.03	Y
89	J045343.50 + 440507.5	8	161.66	+0.18	5127 ± 06	100 ± 18	78 ± 24	2.41 ± 0.42	0.28 ± 0.09	3.17	73	8.54	Y
90	J045355.98 + 443926.5	5	161.24	+0.57	13987 ± 19	104 ± 57	57 ± 76	1.76 ± 0.41	0.13 ± 0.10	1.69	200	9.07	–
91	J045358.02 + 461404.7	2	160.02	+1.57	7124 ± 01	164 ± 03	147 ± 04	5.51 ± 0.49	0.99 ± 0.10	8.27	102	9.38	Y
92	J045408.90 + 444510.6	5	161.19	+0.66	5046 ± 02	132 ± 06	112 ± 08	4.23 ± 0.45	0.71 ± 0.09	6.83	72	8.94	Y
93	J045409.39 + 444601.2	5	161.18	+0.67	9973 ± 02	112 ± 06	99 ± 08	3.78 ± 0.44	0.74 ± 0.08	7.56	142	9.55	Y
94	J045411.00 + 461648.8	2	160.01	+1.63	9011 ± 04	214 ± 12	195 ± 16	3.09 ± 0.43	0.40 ± 0.14	2.91	129	9.19	Y
95	J045414.49 + 450259.2	5	160.97	+0.86	6244 ± 04	180 ± 12	161 ± 16	2.28 ± 0.42	0.41 ± 0.11	3.31	89	8.89	Y
96	J045419.09 + 454741.2	2	160.40	+1.34	10838 ± 03	167 ± 09	132 ± 12	2.66 ± 0.42	0.69 ± 0.08	6.13	155	9.59	Y

TABLE A.1: Visual method galaxy catalogue consisting of 194 detections.

No.	Source ID	Subcube	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	$\log(M_{HI})$	SAC
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
97	J045426.13 + 444958.6	5	161.16	+0.75	5060 ± 02	133 ± 06	104 ± 08	2.69 ± 0.42	0.99 ± 0.06	9.86	72	9.09	Y
98	J045426.23 + 462804.4	2	159.89	+1.78	14681 ± 04	183 ± 12	160 ± 16	1.97 ± 0.41	0.46 ± 0.09	3.66	210	9.68	Y
99	J045428.05 + 453808.4	5	160.54	+1.26	6891 ± 25	101 ± 75	71 ± 100	1.37 ± 0.41	0.08 ± 0.13	0.92	98	8.24	–
100	J045429.99 + 461355.9	2	160.08	+1.64	6706 ± 02	232 ± 06	216 ± 08	4.08 ± 0.45	0.87 ± 0.11	6.01	96	9.28	Y
101	J045431.53 + 444653.5	5	161.21	+0.73	10183 ± 02	201 ± 06	188 ± 08	1.73 ± 0.41	0.78 ± 0.07	5.77	145	9.59	Y
102	J045443.54 + 453933.8	5	160.55	+1.31	4934 ± 01	296 ± 03	228 ± 04	6.89 ± 0.53	5.36 ± 0.06	36.00	70	9.80	Y
103	J045443.67 + 444705.6	5	161.23	+0.76	9907 ± 01	106 ± 03	88 ± 04	5.03 ± 0.47	1.00 ± 0.07	10.81	142	9.67	Y
104	J045444.39 + 462120.8	2	160.01	+1.75	9863 ± 04	302 ± 12	89 ± 16	3.49 ± 0.44	1.17 ± 0.06	12.63	141	9.74	Y
105	J045444.39 + 462120.8	2	160.01	+1.75	9489 ± 05	82 ± 15	43 ± 20	12.04 ± 0.41	0.42 ± 0.05	6.58	136	9.27	Y
106	J045445.56 + 442153.3	8	161.56	+0.50	10494 ± 01	226 ± 03	205 ± 04	3.60 ± 0.44	1.37 ± 0.08	9.74	150	9.86	Y
107	J045456.00 + 463745.1	2	159.82	+1.95	6849 ± 09	130 ± 27	106 ± 36	2.31 ± 0.42	0.20 ± 0.12	1.98	98	8.66	Y
108	J045457.94 + 444649.6	5	161.26	+0.79	10001 ± 07	56 ± 21	44 ± 28	2.00 ± 0.41	0.17 ± 0.08	2.66	143	8.92	Y
109	J045508.07 + 440332.7	8	161.84	+0.36	9766 ± 10	93 ± 30	51 ± 40	1.66 ± 0.41	0.22 ± 0.07	3.13	140	9.00	–
110	J045514.67 + 453833.6	5	160.62	+1.37	3334 ± 04	121 ± 12	101 ± 16	3.37 ± 0.43	0.44 ± 0.10	4.44	48	8.37	Y
111	J045515.30 + 442743.6	8	161.54	+0.63	10044 ± 08	102 ± 24	36 ± 32	2.07 ± 0.41	0.35 ± 0.05	5.89	143	9.23	Y
112	J045518.01 + 442415.7	8	161.59	+0.60	10000 ± 11	65 ± 33	46 ± 44	1.94 ± 0.41	0.14 ± 0.09	2.08	143	8.83	Y
113	J045539.76 + 462253.2	2	160.09	+1.89	9471 ± 06	90 ± 18	78 ± 24	2.23 ± 0.42	0.21 ± 0.10	2.38	135	8.95	Y
114	J045542.03 + 461444.3	2	160.20	+1.81	9790 ± 09	104 ± 27	56 ± 36	2.94 ± 0.43	0.26 ± 0.09	3.46	140	9.07	Y
115	J045544.96 + 444504.8	5	161.37	+0.88	10163 ± 03	182 ± 09	166 ± 12	2.70 ± 0.42	0.44 ± 0.11	3.47	145	9.34	Y
116	J045546.54 + 464912.6	2	159.76	+2.18	4639 ± 26	83 ± 78	54 ± 104	1.80 ± 0.41	0.07 ± 0.13	0.99	66	7.87	Y
117	J045548.03 + 452923.9	5	160.80	+1.35	14102 ± 05	233 ± 15	189 ± 20	2.48 ± 0.42	0.47 ± 0.11	3.49	201	9.66	Y
118	J045550.84 + 440334.4	8	161.92	+0.46	9679 ± 11	92 ± 33	49 ± 44	2.35 ± 0.42	0.21 ± 0.09	2.99	138	8.97	Y
119	J045556.96 + 464601.8	2	159.82	+2.17	6497 ± 13	87 ± 39	50 ± 52	1.11 ± 0.40	0.16 ± 0.07	2.31	93	8.51	Y
120	J045556.96 + 464601.8	2	159.82	+2.17	7089 ± 06	131 ± 18	116 ± 24	2.42 ± 0.42	0.22 ± 0.13	2.10	101	8.73	Y
121	J045557.77 + 461218.5	2	160.26	+1.82	10215 ± 06	92 ± 18	53 ± 24	2.00 ± 0.41	0.38 ± 0.06	5.26	146	9.28	Y
122	J045601.85 + 444510.9	5	161.40	+0.92	10569 ± 05	95 ± 15	69 ± 20	1.13 ± 0.40	0.34 ± 0.05	4.17	151	9.26	–
123	J045602.99 + 452957.0	5	160.82	+1.39	6667 ± 01	248 ± 03	222 ± 04	6.23 ± 0.51	1.85 ± 0.10	12.61	95	9.60	Y
124	J045606.36 + 462548.0	2	160.10	+1.98	9143 ± 04	190 ± 12	168 ± 16	2.25 ± 0.42	0.38 ± 0.11	2.99	131	9.19	Y
125	J045606.53 + 445246.3	5	161.31	+1.01	14079 ± 07	69 ± 21	52 ± 28	1.74 ± 0.41	0.21 ± 0.07	2.94	201	9.30	–
126	J045612.36 + 445031.5	5	161.35	+1.00	13885 ± 12	89 ± 36	65 ± 48	1.23 ± 0.40	0.14 ± 0.09	1.76	198	9.11	–
127	J045614.62 + 455612.1	2	160.50	+1.69	6977 ± 01	157 ± 03	135 ± 04	3.61 ± 0.44	1.12 ± 0.07	9.77	100	9.42	–
128	J045622.25 + 452545.3	5	160.91	+1.39	6697 ± 11	181 ± 33	81 ± 44	1.18 ± 0.40	0.30 ± 0.06	3.43	96	8.82	–

TABLE A.1: Visual method galaxy catalogue consisting of 194 detections.

No.	Source ID	Subcube	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	$\log(M_{HI})$	SAC
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
129	J045626.68 + 443843.3	5	161.53	+0.91	10515 ± 17	146 ± 51	88 ± 68	1.65 ± 0.41	0.16 ± 0.11	1.73	150	8.93	Y
130	J045631.96 + 443147.4	8	161.63	+0.85	10460 ± 08	60 ± 24	28 ± 32	4.15 ± 0.45	0.25 ± 0.08	4.81	149	9.12	Y
131	J045635.65 + 460211.9	2	160.46	+1.80	6661 ± 01	304 ± 03	277 ± 04	3.60 ± 0.44	1.59 ± 0.09	9.68	95	9.53	Y
132	J045647.77 + 444924.3	5	161.43	+1.07	9619 ± 06	172 ± 18	158 ± 24	2.44 ± 0.42	0.20 ± 0.16	1.63	137	8.95	Y
133	J045653.53 + 443537.9	5	161.62	+0.94	10609 ± 09	286 ± 27	220 ± 36	1.92 ± 0.41	0.33 ± 0.13	2.28	152	9.26	Y
134	J045654.08 + 461139.5	2	160.37	+1.94	6888 ± 05	126 ± 15	112 ± 20	1.89 ± 0.41	0.29 ± 0.10	2.74	98	8.82	Y
135	J045702.04 + 444907.3	5	161.46	+1.10	14650 ± 04	245 ± 12	105 ± 16	1.53 ± 0.41	0.93 ± 0.05	9.16	209	9.98	Y
136	J045704.58 + 443407.8	8	161.66	+0.95	14857 ± 26	69 ± 78	54 ± 104	1.92 ± 0.41	0.05 ± 0.16	0.72	212	8.74	Y
137	J045708.37 + 444742.9	5	161.49	+1.10	14578 ± 03	255 ± 09	74 ± 12	1.45 ± 0.41	1.68 ± 0.03	19.86	208	10.24	Y
138	J045714.23 + 445308.9	5	161.43	+1.17	14656 ± 02	247 ± 06	226 ± 08	1.79 ± 0.41	0.74 ± 0.08	4.99	209	9.88	Y
139	J045714.44 + 451221.3	5	161.18	+1.37	6266 ± 04	237 ± 12	184 ± 16	2.48 ± 0.42	0.66 ± 0.09	4.95	90	9.10	Y
140	J045718.48 + 453727.6	5	160.86	+1.64	8275 ± 39	234 ± 17	173 ± 156	1.26 ± 0.40	0.07 ± 0.20	0.54	118	8.36	–
141	J045719.11 + 461942.5	2	160.31	+2.08	7044 ± 14	57 ± 42	32 ± 56	1.87 ± 0.41	0.12 ± 0.08	2.21	101	8.47	–
142	J045721.22 + 454140.5	5	160.81	+1.69	3403 ± 01	119 ± 03	101 ± 04	6.04 ± 0.50	2.80 ± 0.05	28.29	49	9.19	Y
143	J045723.71 + 463407.8	2	160.13	+2.24	7045 ± 02	213 ± 06	195 ± 08	3.27 ± 0.43	0.61 ± 0.12	4.43	101	9.16	Y
144	J045725.24 + 440232.9	8	162.11	+0.67	9684 ± 06	148 ± 18	116 ± 24	2.69 ± 0.42	0.35 ± 0.11	3.35	138	9.21	Y
145	J045733.21 + 444855.5	5	161.52	+1.17	14404 ± 05	251 ± 15	223 ± 20	1.90 ± 0.41	0.38 ± 0.12	2.59	206	9.58	Y
146	J045733.43 + 453410.6	5	160.93	+1.64	8986 ± 08	46 ± 24	40 ± 32	1.34 ± 0.41	0.11 ± 0.08	1.83	128	8.65	–
147	J045740.93 + 453427.2	5	160.94	+1.66	6269 ± 05	100 ± 15	73 ± 20	3.25 ± 0.43	0.35 ± 0.09	4.20	90	8.83	Y
148	J045745.40 + 460714.2	2	160.52	+2.01	6997 ± 00	397 ± 00	378 ± 00	2.12 ± 0.41	3.13 ± 0.06	16.33	100	9.87	Y
149	J045748.40 + 455655.9	2	160.66	+1.91	14554 ± 06	182 ± 18	86 ± 24	1.68 ± 0.41	0.60 ± 0.06	6.52	208	9.78	Y
150	J045749.98 + 455537.1	1	160.68	+1.90	10808 ± 01	235 ± 03	191 ± 04	5.80 ± 0.49	2.18 ± 0.08	16.02	154	10.09	Y
151	J045751.10 + 455718.3	1	160.66	+1.92	14567 ± 21	145 ± 63	83 ± 84	1.68 ± 0.41	0.13 ± 0.12	1.45	208	9.12	Y
152	J045751.12 + 461930.2	1	160.37	+2.15	6842 ± 01	115 ± 03	101 ± 04	4.24 ± 0.45	1.55 ± 0.06	15.61	98	9.54	Y
153	J045757.35 + 462846.3	1	160.26	+2.26	6973 ± 07	275 ± 21	71 ± 28	2.87 ± 0.43	0.75 ± 0.06	9.00	100	9.24	Y
154	J045804.01 + 445533.5	4	161.49	+1.31	9462 ± 27	146 ± 81	21 ± 108	1.97 ± 0.41	0.14 ± 0.06	3.19	135	8.79	Y
155	J045807.00 + 454801.2	1	160.81	+1.86	6555 ± 02	224 ± 06	192 ± 08	4.09 ± 0.45	0.91 ± 0.10	6.65	94	9.27	Y
156	J045807.97 + 450517.0	4	161.37	+1.42	3525 ± 00	192 ± 00	175 ± 00	9.63 ± 0.63	4.67 ± 0.07	35.84	50	9.45	Y
157	J045808.53 + 450608.5	4	161.36	+1.43	10061 ± 03	272 ± 09	258 ± 12	1.88 ± 0.41	0.44 ± 0.12	2.80	144	9.33	Y
158	J045811.74 + 460226.5	1	160.63	+2.02	7121 ± 01	221 ± 03	207 ± 04	4.46 ± 0.46	1.13 ± 0.10	7.94	102	9.44	Y
159	J045827.09 + 441339.5	7	162.08	+0.93	14591 ± 04	334 ± 12	314 ± 16	1.26 ± 0.40	0.44 ± 0.11	2.54	208	9.66	Y
160	J045834.88 + 453113.6	4	161.08	+1.75	5022 ± 03	150 ± 09	131 ± 12	3.24 ± 0.43	0.57 ± 0.10	5.08	72	8.84	Y

TABLE A.1: Visual method galaxy catalogue consisting of 194 detections.

No.	Source ID	Subcube	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	$\log(M_{HI})$	SAC
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
161	J045837.19 + 461809.2	1	160.47	+2.24	7034 ± 01	201 ± 03	181 ± 04	6.23 ± 0.51	1.48 ± 0.10	11.13	100	9.55	Y
162	J045842.28 + 455641.9	1	160.76	+2.03	6753 ± 01	245 ± 03	216 ± 04	3.61 ± 0.44	1.29 ± 0.09	8.91	96	9.45	Y
163	J045846.46 + 441406.5	7	162.11	+0.98	6870 ± 01	192 ± 03	161 ± 04	9.10 ± 0.61	2.36 ± 0.09	18.91	98	9.73	Y
164	J045851.06 + 445344.1	4	161.60	+1.40	5191 ± 08	136 ± 24	113 ± 32	2.14 ± 0.41	0.20 ± 0.12	1.96	74	8.42	Y
165	J045855.91 + 454752.6	1	160.90	+1.97	5072 ± 07	212 ± 21	77 ± 28	3.80 ± 0.44	0.60 ± 0.08	6.97	72	8.87	Y
166	J045906.48 + 440442.8	7	162.27	+0.93	6767 ± 01	119 ± 03	91 ± 04	7.36 ± 0.54	2.52 ± 0.06	26.80	97	9.74	Y
167	J045908.43 + 451600.7	4	161.34	+1.67	4892 ± 06	195 ± 18	141 ± 24	2.37 ± 0.42	0.42 ± 0.10	3.62	70	8.69	Y
168	J045913.08 + 463039.6	1	160.37	+2.45	7048 ± 07	103 ± 21	86 ± 28	1.50 ± 0.41	0.21 ± 0.09	2.29	101	8.70	Y
169	J045921.24 + 441946.8	7	162.10	+1.12	10626 ± 02	186 ± 06	160 ± 08	3.02 ± 0.43	0.99 ± 0.08	7.92	152	9.73	Y
170	J045930.09 + 441844.2	7	162.13	+1.13	9888 ± 02	274 ± 06	259 ± 08	2.23 ± 0.42	0.66 ± 0.11	4.16	141	9.49	Y
171	J045931.11 + 455242.0	1	160.90	+2.10	11952 ± 10	227 ± 30	103 ± 40	2.60 ± 0.42	0.39 ± 0.09	3.88	171	9.43	Y
172	J045931.11 + 455242.0	1	160.90	+2.10	6683 ± 14	199 ± 42	126 ± 56	1.46 ± 0.41	0.21 ± 0.11	1.89	95	8.65	Y
173	J045936.73 + 445702.8	4	161.64	+1.54	10176 ± 02	210 ± 06	189 ± 08	3.80 ± 0.44	0.72 ± 0.11	5.35	145	9.56	Y
174	J045939.51 + 452927.4	4	161.22	+1.88	7008 ± 01	160 ± 03	145 ± 04	4.25 ± 0.45	2.42 ± 0.06	20.37	100	9.76	Y
175	J045945.04 + 455133.2	1	160.94	+2.12	13304 ± 08	213 ± 24	156 ± 32	2.06 ± 0.41	0.34 ± 0.11	2.79	190	9.47	Y
176	J045946.27 + 462125.8	1	160.55	+2.43	5163 ± 08	90 ± 24	42 ± 32	1.21 ± 0.40	0.30 ± 0.05	4.67	74	8.58	Y
177	J045950.87 + 453146.8	4	161.21	+1.93	10850 ± 15	131 ± 45	92 ± 60	1.83 ± 0.41	0.14 ± 0.12	1.52	155	8.91	Y
178	J045955.41 + 445938.3	4	161.64	+1.61	10191 ± 09	124 ± 27	67 ± 36	1.64 ± 0.41	0.28 ± 0.07	3.48	146	9.15	Y
179	J045957.14 + 440545.8	7	162.35	+1.06	5669 ± 04	204 ± 12	171 ± 16	2.19 ± 0.41	0.51 ± 0.10	3.92	81	8.89	Y
180	J050008.44 + 452200.4	4	161.37	+1.87	4894 ± 03	148 ± 09	132 ± 12	3.22 ± 0.43	0.43 ± 0.11	3.79	70	8.69	Y
181	J050039.53 + 453135.6	4	161.30	+2.04	3368 ± 02	177 ± 06	150 ± 08	3.12 ± 0.43	0.85 ± 0.08	7.05	48	8.67	Y
182	J050044.28 + 454509.9	4	161.13	+2.19	3423 ± 10	47 ± 30	32 ± 40	2.71 ± 0.42	0.13 ± 0.09	2.37	49	7.87	Y
183	J050117.77 + 454725.8	1	161.16	+2.29	3349 ± 01	175 ± 03	153 ± 04	6.58 ± 0.52	1.22 ± 0.10	9.98	48	8.82	Y
184	J050120.41 + 442244.1	7	162.28	+1.43	0688 ± 12	109 ± 36	86 ± 48	1.63 ± 0.41	0.14 ± 0.11	1.52	153	8.88	Y
185	J050124.23 + 452229.9	4	161.50	+2.05	6092 ± 13	144 ± 39	106 ± 52	1.92 ± 0.41	0.17 ± 0.13	1.63	87	8.47	–
186	J050124.51 + 441456.9	7	162.39	+1.36	8833 ± 09	107 ± 27	93 ± 36	1.85 ± 0.41	0.14 ± 0.12	1.50	126	8.73	Y
187	J050217.10 + 455440.2	1	161.17	+2.50	13362 ± 05	365 ± 15	335 ± 20	1.47 ± 0.41	0.39 ± 0.13	2.18	191	9.53	Y
188	J050228.64 + 451047.7	4	161.77	+2.08	8284 ± 06	317 ± 18	294 ± 24	1.25 ± 0.40	0.28 ± 0.13	1.63	118	8.96	Y
189	J050237.29 + 461144.8	1	160.98	+2.72	9635 ± 01	127 ± 03	109 ± 04	3.22 ± 0.43	1.02 ± 0.07	9.91	138	9.66	Y
190	J050252.10 + 451614.9	4	161.74	+2.19	14476 ± 05	299 ± 15	278 ± 20	1.57 ± 0.41	0.30 ± 0.14	1.81	207	9.48	Y
191	J050256.46 + 451217.0	4	161.80	+2.16	3213 ± 01	145 ± 03	126 ± 04	5.79 ± 0.49	1.22 ± 0.09	10.99	46	8.78	Y
192	J050319.61 + 450104.4	4	161.99	+2.10	8490 ± 03	117 ± 09	95 ± 12	2.87 ± 0.42	0.48 ± 0.09	4.99	121	9.22	Y

TABLE A.1: Visual method galaxy catalogue consisting of 194 detections.

No.	**Source ID**	**Subcube**	$\boldsymbol{\ell}$	**b**	$\mathbf{V_{hel}}$	$\mathbf{W_{20}}$	$\mathbf{W_{50}}$	$\mathbf{S_{peak}}$	$\mathbf{S_{int}}$	SNR	D	$\mathbf{\log(M_{HI})}$	SAC
			deg	deg	$\mathrm{km\,s^{-1}}$	$\mathrm{km\,s^{-1}}$	$\mathrm{km\,s^{-1}}$	$\mathrm{mJy\,beam^{-1}}$	$\mathrm{Jy\,km\,s^{-1}}$	$--$	Mpc	$\mathrm{M_\odot}$	$--$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
193	J050329.29 + 454611.4	1	161.41	+2.58	5476 ± 03	286 ± 09	241 ± 12	2.76 ± 0.42	0.87 ± 0.10	5.71	78	9.10	Y
194	J050333.73 + 442813.8	7	162.45	+1.80	7061 ± 22	207 ± 66	$77. \pm 88$	1.50 ± 0.41	0.18 ± 0.09	2.10	101	8.64	Y

Appendix B

SoFiA catalogue

TABLE B.1: SoFiA galaxy catalogue consisting of 131 detections.

No.	Source ID	SC	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	Mass	CM
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	J044521.33 + 454433.8	6	159.44	+0.11	5099 ± 01	294 ± 03	277 ± 04	3.20 ± 0.43	0.93 ± 0.11	5.64	73	9.06	Y
2	J044542.93 + 442103.0	9	160.54	−0.75	10687 ± 08	432 ± 24	170 ± 32	2.40 ± 0.42	0.30 ± 0.13	2.36	153	9.22	Y
3	J044556.86 + 450740.7	6	159.97	−0.21	6776 ± 01	86 ± 03	70 ± 04	2.80 ± 0.42	0.34 ± 0.09	4.15	97	8.88	Y
4	J044602.28 + 443415.4	9	160.41	−0.56	6252 ± 00	110 ± 00	97 ± 00	4.60 ± 0.46	2.10 ± 0.05	21.65	89	9.60	Y
5	J044632.73 + 452147.5	6	159.86	+0.02	9697 ± 03	332 ± 09	250 ± 12	3.60 ± 0.44	1.04 ± 0.10	6.65	139	9.67	Y
6	J044635.79 + 441908.8	9	160.66	−0.65	9968 ± 01	126 ± 03	97 ± 04	2.40 ± 0.42	0.34 ± 0.09	3.53	142	9.21	Y
7	J044644.14 + 442002.3	9	160.67	−0.62	5619 ± 01	193 ± 03	175 ± 04	3.60 ± 0.44	0.66 ± 0.11	5.08	80	9.00	Y
8	J044644.75 + 444735.4	6	160.32	−0.32	5687 ± 01	154 ± 03	137 ± 04	2.90 ± 0.43	0.48 ± 0.10	4.16	81	8.87	Y
9	J044652.85 + 440900.6	9	160.82	−0.72	7471 ± 02	156 ± 06	112 ± 08	2.50 ± 0.42	0.32 ± 0.11	3.09	107	8.94	Y
10	J044700.42 + 442441.1	9	160.64	−0.54	10710 ± 03	520 ± 09	309 ± 12	5.10 ± 0.47	0.85 ± 0.15	4.89	153	9.67	Y
11	J044706.89 + 453448.6	6	159.76	+0.24	4962 ± 00	212 ± 00	200 ± 00	5.10 ± 0.47	1.32 ± 0.10	9.48	71	9.19	Y
12	J044719.79 + 435437.3	9	161.06	−0.82	10038 ± 01	199 ± 03	186 ± 04	2.40 ± 0.42	0.40 ± 0.12	2.95	143	9.28	Y
13	J044728.24 + 444713.2	6	160.41	−0.23	7272 ± 01	152 ± 03	125 ± 04	2.90 ± 0.43	0.76 ± 0.08	6.94	104	9.29	Y
14	J044727.41 + 445341.1	6	160.32	−0.16	5493 ± 01	70 ± 03	49 ± 04	3.20 ± 0.43	0.32 ± 0.08	4.59	78	8.66	Y
15	J044744.05 + 441944.8	9	160.79	−0.49	4847 ± 01	142 ± 03	110 ± 04	4.00 ± 0.45	0.84 ± 0.08	8.17	69	8.98	Y
16	J044809.30 + 443633.4	6	160.62	−0.25	9846 ± 01	103 ± 03	81 ± 04	2.60 ± 0.42	0.23 ± 0.11	2.61	141	9.03	Y
17	J044838.31 + 454505.1	6	159.80	+0.55	5722 ± 01	166 ± 03	125 ± 04	2.90 ± 0.43	0.66 ± 0.08	6.02	82	9.02	Y
18	J044840.48 + 452428.5	6	160.07	+0.33	10930 ± 01	165 ± 03	153 ± 04	3.00 ± 0.43	0.42 ± 0.12	3.46	156	9.38	Y
19	J044844.84 + 435610.4	9	161.20	−0.60	3375 ± 01	37 ± 03	24 ± 04	3.60 ± 0.44	0.74 ± 0.04	15.30	48	8.61	Y
20	J044901.29 + 441926.4	9	160.94	−0.32	14494 ± 03	298 ± 09	140 ± 12	2.20 ± 0.41	0.36 ± 0.10	3.05	207	9.56	Y
21	J044912.99 + 442510.1	9	160.89	−0.23	8284 ± 02	248 ± 06	197 ± 08	3.10 ± 0.43	0.89 ± 0.09	6.43	118	9.47	Y
22	J044934.83 + 451032.7	6	160.35	+0.31	6627 ± 01	78 ± 03	40 ± 04	3.40 ± 0.43	0.28 ± 0.08	4.51	95	8.77	Y
23	J044935.92 + 445251.7	6	160.58	+0.12	3956 ± 01	70 ± 03	37 ± 04	2.60 ± 0.42	0.18 ± 0.08	3.01	57	8.13	Y
24	J045002.35 + 452643.6	6	160.19	+0.54	9588 ± 01	121 ± 03	109 ± 04	3.30 ± 0.43	0.51 ± 0.10	4.94	137	9.35	Y
25	J045002.55 + 444902.6	6	160.68	+0.14	3438 ± 02	69 ± 06	34 ± 08	2.70 ± 0.42	0.57 ± 0.04	10.00	49	8.51	Y
26	J045005.31 + 453020.4	6	160.15	+0.59	9676 ± 04	298 ± 12	277 ± 16	4.50 ± 0.46	0.22 ± 0.27	1.35	138	9.00	Y
27	J045005.38 + 450301.0	6	160.50	+0.29	14882 ± 07	145 ± 21	64 ± 28	3.00 ± 0.43	0.11 ± 0.15	1.37	213	9.06	Y
28	J045013.15 + 440832.8	9	161.22	−0.27	5528 ± 01	136 ± 03	110 ± 04	4.70 ± 0.46	0.83 ± 0.09	8.04	79	9.09	Y
29	J045025.97 + 451150.2	6	160.43	+0.43	4627 ± 03	170 ± 09	96 ± 12	4.30 ± 0.45	0.87 ± 0.08	9.00	66	8.95	Y
30	J045035.02 + 442221.1	9	161.08	−0.07	4303 ± 01	50 ± 03	32 ± 04	3.30 ± 0.43	0.13 ± 0.10	2.29	61	8.06	Y
31	J045104.77 + 443523.0	9	160.97	+0.13	5043 ± 01	127 ± 03	67 ± 04	7.40 ± 0.54	1.90 ± 0.06	23.56	72	9.37	Y
32	J045105.23 + 443545.0	6	160.97	+0.14	5009 ± 00	188 ± 00	170 ± 00	7.40 ± 0.54	3.11 ± 0.07	24.17	72	9.57	Y
33	J045137.07 + 461913.7	2	159.70	+1.31	7002 ± 02	115 ± 06	77 ± 08	2.40 ± 0.42	0.37 ± 0.08	4.24	100	8.94	Y

TABLE B.1: SoFiA galaxy catalogue consisting of 131 detections.

No.	Source ID	SC	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	Mass	CM
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	$--$	Mpc	$M_\odot$	$--$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
34	J045136.22+452437.2	5	160.40	+0.73	9623 ± 01	43 ± 03	31 ± 04	3.50 ± 0.44	0.20 ± 0.08	3.61	137	8.95	Y
35	J045138.49+443604.9	5	161.03	+0.22	5044 ± 00	258 ± 00	219 ± 00	8.60 ± 0.59	9.77 ± 0.05	66.97	72	10.08	Y
36	J045208.86+434930.1	8	161.68	−0.20	7262 ± 01	322 ± 03	304 ± 04	5.10 ± 0.47	0.91 ± 0.15	5.31	104	9.36	Y
37	J045217.62+450844.9	5	160.68	+0.65	6643 ± 02	160 ± 06	118 ± 08	3.10 ± 0.43	0.35 ± 0.12	3.23	95	8.87	Y
38	J045220.63+454136.1	5	160.26	+1.01	6429 ± 00	182 ± 00	170 ± 00	4.10 ± 0.45	0.78 ± 0.11	6.04	92	9.19	Y
39	J045223.77+452743.9	5	160.44	+0.87	5946 ± 01	95 ± 03	84 ± 04	3.30 ± 0.43	0.33 ± 0.10	3.68	85	8.75	Y
40	J045227.25+444541.1	5	160.99	+0.43	5115 ± 01	105 ± 03	90 ± 04	2.40 ± 0.42	0.34 ± 0.09	3.66	73	8.63	Y
41	J045231.96+461224.7	2	159.88	+1.36	7063 ± 02	193 ± 06	110 ± 08	3.50 ± 0.44	0.54 ± 0.10	5.19	101	9.11	Y
42	J045256.01+460256.6	2	160.05	+1.31	7649 ± 03	142 ± 09	89 ± 12	2.30 ± 0.42	0.28 ± 0.10	3.00	109	8.89	Y
43	J045358.15+461405.0	2	160.02	+1.57	7113 ± 00	160 ± 00	141 ± 00	5.50 ± 0.49	1.07 ± 0.10	9.16	102	9.42	Y
44	J045408.61+444522.2	5	161.19	+0.66	5026 ± 01	131 ± 03	112 ± 04	4.20 ± 0.45	0.85 ± 0.08	8.19	72	9.02	Y
45	J045410.94+444559.6	5	161.18	+0.67	9969 ± 01	113 ± 03	97 ± 04	3.80 ± 0.44	0.45 ± 0.10	4.64	142	9.33	Y
46	J045414.10+461632.4	2	160.02	+1.63	8989 ± 05	310 ± 15	197 ± 20	3.10 ± 0.43	0.27 ± 0.17	1.93	128	9.02	Y
47	J045415.14+450316.2	5	160.97	+0.86	6229 ± 02	187 ± 06	140 ± 08	2.30 ± 0.42	0.48 ± 0.09	4.15	89	8.96	Y
48	J045420.59+454724.3	2	160.41	+1.34	10816 ± 01	151 ± 03	120 ± 04	2.70 ± 0.42	0.51 ± 0.09	4.70	155	9.46	Y
49	J045426.84+444941.3	5	161.16	+0.75	5043 ± 01	122 ± 03	104 ± 04	2.70 ± 0.42	0.53 ± 0.08	5.23	72	8.81	Y
50	J045431.23+461343.5	2	160.09	+1.64	6674 ± 01	222 ± 03	211 ± 04	4.60 ± 0.46	0.54 ± 0.15	3.79	95	9.07	Y
51	J045443.31+462110.9	2	160.01	+1.75	9922 ± 14	228 ± 42	89 ± 56	4.00 ± 0.45	0.16 ± 0.17	1.76	142	8.89	Y
52	J045443.57+444708.0	5	161.23	+0.76	9896 ± 00	105 ± 00	87 ± 00	5.00 ± 0.47	1.20 ± 0.07	13.03	141	9.75	Y
53	J045445.40+453929.0	5	160.55	+1.31	4864 ± 01	288 ± 03	224 ± 04	8.40 ± 0.58	3.73 ± 0.08	25.29	69	9.63	Y
54	J045446.86+442140.1	8	161.56	+0.50	10478 ± 01	225 ± 03	205 ± 04	4.20 ± 0.45	1.59 ± 0.08	11.27	150	9.92	Y
55	J045457.53+463738.0	2	159.82	+1.95	6839 ± 04	168 ± 12	51 ± 16	2.30 ± 0.42	0.26 ± 0.08	3.70	98	8.77	Y
56	J045514.27+453815.4	5	160.62	+1.37	3319 ± 01	117 ± 03	98 ± 04	3.40 ± 0.43	0.51 ± 0.09	5.27	47	8.44	Y
57	J045543.01+461422.0	2	160.21	+1.81	9762 ± 05	242 ± 15	61 ± 20	2.90 ± 0.43	0.32 ± 0.08	4.22	139	9.17	Y
58	J045546.35+444512.2	5	161.37	+0.88	10166 ± 04	299 ± 12	184 ± 16	2.70 ± 0.42	0.59 ± 0.10	4.43	145	9.47	Y
59	J045547.40+452914.4	5	160.80	+1.35	14076 ± 01	192 ± 03	182 ± 04	2.50 ± 0.42	0.43 ± 0.12	3.25	201	9.62	Y
60	J045604.57+452940.8	5	160.83	+1.39	6671 ± 00	237 ± 00	213 ± 00	6.20 ± 0.51	1.85 ± 0.10	12.87	95	9.60	Y
61	J045616.17+455602.6	2	160.50	+1.69	6961 ± 01	148 ± 03	131 ± 04	3.60 ± 0.44	0.62 ± 0.10	5.51	99	9.16	Y
62	J045633.53+443126.4	8	161.64	+0.85	10449 ± 02	63 ± 06	29 ± 08	4.20 ± 0.45	0.30 ± 0.07	5.59	149	9.19	Y
63	J045635.14+460228.3	2	160.46	+1.80	6577 ± 01	303 ± 03	278 ± 04	4.70 ± 0.46	1.05 ± 0.13	6.36	94	9.34	Y
64	J045647.56+444923.9	5	161.43	+1.07	9654 ± 13	190 ± 39	148 ± 52	2.40 ± 0.42	0.05 ± 0.30	0.41	138	8.35	Y
65	J045709.20+444800.9	5	161.49	+1.11	14609 ± 09	246 ± 27	47 ± 36	2.60 ± 0.42	0.35 ± 0.07	5.25	209	9.56	Y
66	J045714.08+451229.4	5	161.18	+1.37	6301 ± 04	233 ± 12	181 ± 16	3.10 ± 0.43	0.30 ± 0.15	2.25	90	8.76	Y

TABLE B.1: SoFiA galaxy catalogue consisting of 131 detections.

No.	Source ID	SC	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	Mass	CM
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	$--$	Mpc	$M_\odot$	$--$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
67	J045720.95 + 454138.7	5	160.81	+1.69	3397 ± 00	117 ± 00	98 ± 00	6.00 ± 0.50	2.46 ± 0.06	25.19	49	9.14	Y
68	J045723.97 + 463356.7	2	160.13	+2.24	7037 ± 01	217 ± 03	203 ± 04	3.30 ± 0.43	0.74 ± 0.11	5.24	101	9.24	Y
69	J045742.81 + 453422.3	5	160.94	+1.66	6251 ± 02	177 ± 06	77 ± 08	3.30 ± 0.43	0.51 ± 0.08	5.94	89	8.99	Y
70	J045749.37 + 460707.6	1	160.53	+2.02	7100 ± 05	300 ± 15	77 ± 20	5.40 ± 0.48	0.56 ± 0.10	6.53	101	9.14	Y
71	J045750.67 + 461921.7	2	160.37	+2.15	6829 ± 01	231 ± 03	103 ± 04	4.20 ± 0.45	1.26 ± 0.07	12.61	98	9.45	Y
72	J045801.11 + 462831.6	1	160.27	+2.27	7025 ± 02	164 ± 06	72 ± 08	2.90 ± 0.43	0.55 ± 0.07	6.63	100	9.12	Y
73	J045808.53 + 450525.5	4	161.37	+1.42	3547 ± 00	190 ± 00	173 ± 00	12.00 ± 0.72	4.06 ± 0.08	31.29	51	9.39	Y
74	J045808.69 + 454734.1	4	160.82	+1.86	6526 ± 01	194 ± 03	181 ± 04	4.10 ± 0.45	0.67 ± 0.12	5.02	93	9.13	Y
75	J045810.82 + 460230.4	1	160.63	+2.02	7140 ± 01	215 ± 03	202 ± 04	4.50 ± 0.46	0.71 ± 0.13	5.05	102	9.24	Y
76	J045825.63 + 462441.8	1	160.36	+2.28	14859 ± 03	65 ± 09	21 ± 12	3.50 ± 0.44	0.11 ± 0.09	2.37	212	9.06	Y
77	J045835.06 + 453111.6	4	161.08	+1.75	5014 ± 01	153 ± 03	134 ± 04	3.20 ± 0.43	0.64 ± 0.09	5.62	72	8.89	Y
78	J045838.01 + 461745.7	1	160.48	+2.24	7007 ± 01	195 ± 03	162 ± 04	6.20 ± 0.51	1.27 ± 0.10	10.11	100	9.48	Y
79	J045842.93 + 455647.2	1	160.76	+2.03	6732 ± 02	238 ± 06	206 ± 08	3.60 ± 0.44	0.41 ± 0.15	2.93	96	8.96	Y
80	J045848.44 + 441347.1	7	162.12	+0.98	6869 ± 01	187 ± 03	160 ± 04	9.10 ± 0.61	1.31 ± 0.12	10.54	98	9.47	Y
81	J045855.72 + 454809.6	1	160.90	+1.97	5059 ± 02	164 ± 06	78 ± 08	3.80 ± 0.44	0.53 ± 0.09	6.06	72	8.81	Y
82	J045908.28 + 440449.9	7	162.27	+0.94	6741 ± 00	118 ± 00	91 ± 00	7.40 ± 0.54	2.69 ± 0.06	28.63	96	9.77	Y
83	J045919.45 + 441941.2	7	162.10	+1.11	10568 ± 03	191 ± 09	175 ± 12	3.00 ± 0.43	0.14 ± 0.22	1.09	151	8.88	Y
84	J045932.74 + 441841.1	7	162.14	+1.14	9888 ± 01	280 ± 03	265 ± 04	3.90 ± 0.45	0.66 ± 0.14	4.13	141	9.49	Y
85	J045936.59 + 445711.4	4	161.64	+1.54	10165 ± 01	209 ± 03	190 ± 04	3.80 ± 0.44	0.83 ± 0.10	6.13	145	9.62	Y
86	J050039.76 + 453120.5	4	161.30	+2.04	3358 ± 01	183 ± 03	147 ± 04	3.10 ± 0.43	0.91 ± 0.08	7.60	48	8.69	Y
87	J050045.75 + 454454.6	4	161.14	+2.19	3412 ± 01	50 ± 03	34 ± 04	2.70 ± 0.42	0.16 ± 0.09	2.77	49	7.95	Y
88	J045752.60 + 455542.1	1	160.68	+1.91	10737 ± 01	223 ± 03	192 ± 04	5.80 ± 0.49	1.12 ± 0.11	8.18	153	9.79	Y
89	J045931.09 + 455250.6	1	160.90	+2.10	11943 ± 02	344 ± 06	236 ± 08	2.60 ± 0.42	0.62 ± 0.11	4.07	171	9.63	Y
90	J045941.72 + 452910.9	4	161.23	+1.88	7006 ± 01	157 ± 03	143 ± 04	4.60 ± 0.46	1.27 ± 0.08	10.76	100	9.48	Y
91	J050007.94 + 452211.6	4	161.37	+1.87	4885 ± 01	147 ± 03	131 ± 04	3.20 ± 0.43	0.48 ± 0.11	4.21	70	8.74	Y
92	J050116.90 + 454729.3	4	161.16	+2.29	3336 ± 01	168 ± 03	145 ± 04	6.60 ± 0.52	1.26 ± 0.10	10.61	48	8.83	Y
93	J050237.32 + 461131.1	1	160.98	+2.72	9623 ± 01	119 ± 03	105 ± 04	3.20 ± 0.43	0.81 ± 0.07	8.04	137	9.56	Y
94	J050256.91 + 451232.1	4	161.80	+2.16	3243 ± 00	145 ± 00	124 ± 00	5.80 ± 0.49	1.43 ± 0.08	13.07	46	8.86	Y
95	J044600.22 + 442444.1	9	160.52	−0.67	13017 ± 06	90 ± 18	66 ± 24	6.20 ± 0.51	0.08 ± 0.25	1.03	186	8.83	–
96	J044602.52 + 443427.9	6	160.40	−0.56	6247 ± 00	112 ± 00	97 ± 00	4.60 ± 0.46	2.69 ± 0.05	27.77	89	9.70	–
97	J044736.90 + 444754.1	6	160.41	−0.20	9756 ± 07	133 ± 21	55 ± 28	1.70 ± 0.41	0.13 ± 0.10	1.77	139	8.77	–
98	J044817.40 + 445346.4	6	160.42	−0.05	13947 ± 23	151 ± 69	117 ± 92	1.40 ± 0.41	0.04 ± 0.23	0.36	199	8.55	–
99	J044819.56 + 444647.3	6	160.51	−0.12	16507 ± 04	65 ± 12	36 ± 16	1.50 ± 0.41	0.11 ± 0.08	1.87	236	9.16	–

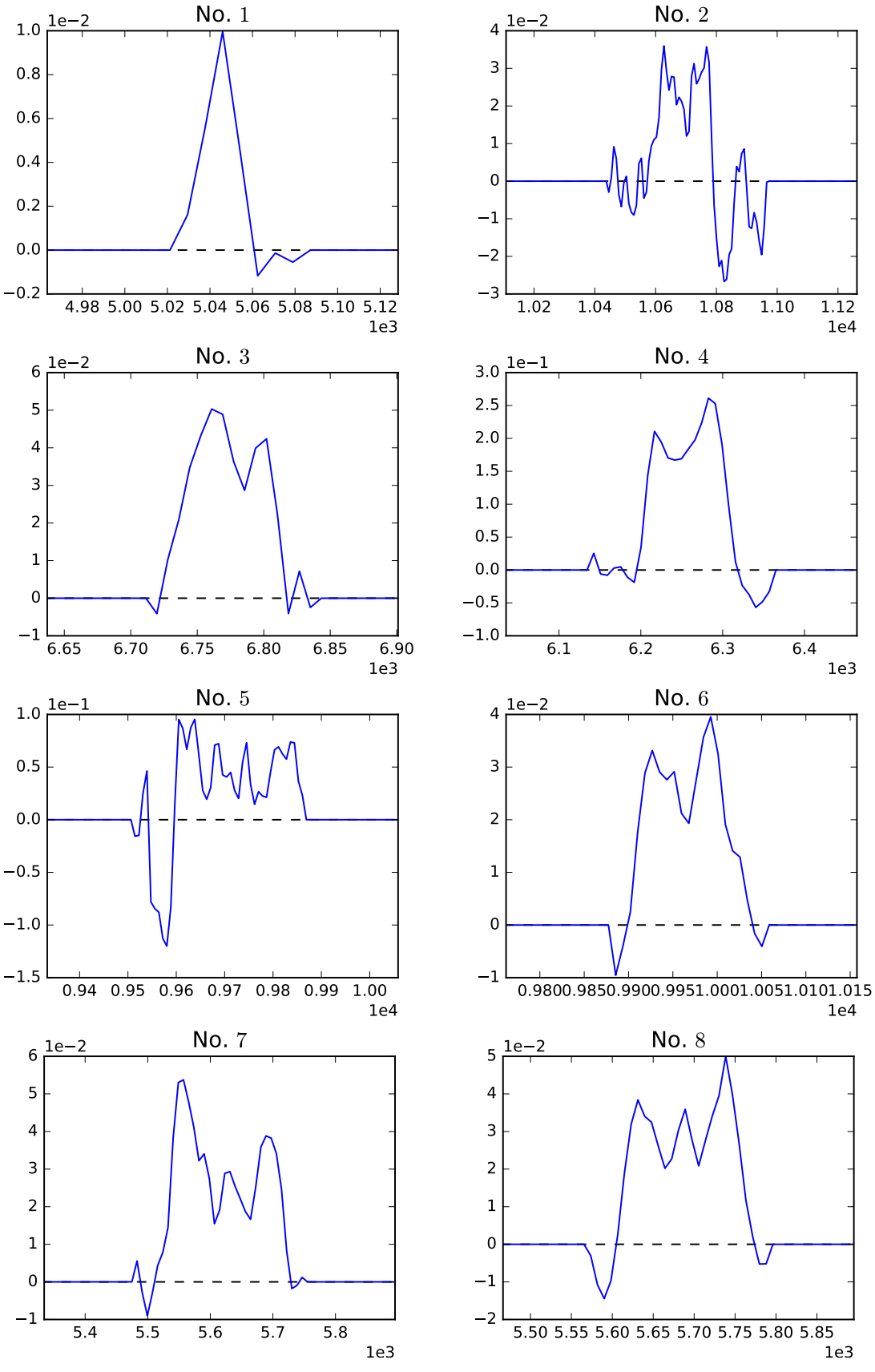
TABLE B.1: SoFiA galaxy catalogue consisting of 131 detections.

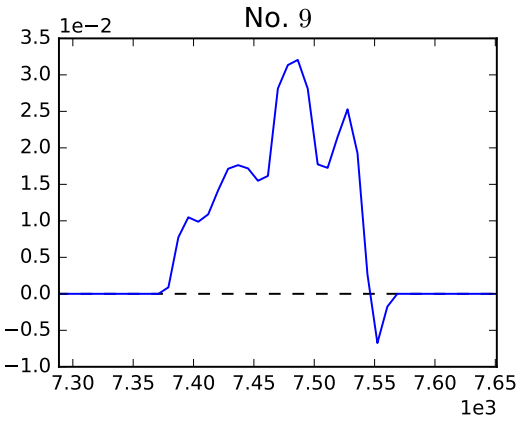
No.	Source ID	SC	ℓ	b	V_{hel}	W_{20}	W_{50}	S_{peak}	S_{int}	SNR	D	Mass	CM
			deg	deg	$km\,s^{-1}$	$km\,s^{-1}$	$km\,s^{-1}$	$mJy\,beam^{-1}$	$Jy\,km\,s^{-1}$	−−	Mpc	$M_\odot$	−−
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
100	J044841.48 + 455316.0	3	159.70	+0.64	12370 ± 15	1191 ± 477	374 ± 636	1.90 ± 0.41	0.08 ± 0.34	0.43	177	8.78	–
101	J044845.25 + 441450.9	9	160.97	−0.40	6606 ± 16	324 ± 48	144. ± 64	1.20 ± 0.40	0.12 ± 0.14	1.02	94	8.40	–
102	J044902.56 + 450145.4	6	160.40	+0.14	12726 ± 23	635 ± 69	375 ± 92	2.40 ± 0.42	0.19 ± 0.25	0.99	182	9.17	–
103	J044902.60 + 450145.6	6	160.40	+0.14	11936 ± 02	144 ± 06	93 ± 08	2.50 ± 0.42	0.68 ± 0.07	7.16	171	9.67	–
104	J044903.17 + 450144.6	6	160.40	+0.14	6795 ± 28	5001 ± 84	143. ± 112	2.60 ± 0.42	0.63 ± 0.09	5.33	97	9.15	–
105	J044943.90 + 460245.3	3	159.70	+0.88	12743 ± 04	656 ± 12	251 ± 16	8.20 ± 0.57	0.65 ± 0.20	4.15	182	9.70	–
106	J044945.22 + 460234.0	3	159.70	+0.88	9466 ± 01	123 ± 03	81 ± 04	18.70 ± 1.02	0.43 ± 0.21	4.88	135	9.27	–
107	J044945.79 + 460144.2	3	159.72	+0.88	16467 ± 02	163 ± 06	122 ± 08	15.40 ± 0.87	0.82 ± 0.17	7.51	235	10.03	–
108	J045039.23 + 434807.5	9	161.53	−0.43	10646 ± 02	61 ± 06	54 ± 08	1.20 ± 0.40	0.09 ± 0.10	1.20	152	8.68	–
109	J045052.33 + 450141.9	6	160.61	+0.39	11934 ± 08	47 ± 24	26 ± 32	1.40 ± 0.41	0.03 ± 0.13	0.52	170	8.25	–
110	J045108.28 + 443522.7	8	160.98	+0.14	5034 ± 00	148 ± 00	123. ± 00	7.40 ± 0.54	2.78 ± 0.07	25.42	72	9.53	–
111	J045143.75 + 441502.4	8	161.31	+0.01	16471 ± 07	92 ± 21	39 ± 28	2.30 ± 0.42	0.08 ± 0.12	1.25	235	9.00	–
112	J045145.98 + 443513.1	8	161.05	+0.23	5134 ± 01	130 ± 03	104 ± 04	6.20 ± 0.51	2.23 ± 0.06	22.16	73	9.45	–
113	J045207.71 + 451535.2	5	160.57	+0.70	15933 ± 02	105 ± 06	94 ± 08	2.40 ± 0.42	0.47 ± 0.08	4.93	228	9.76	–
114	J045229.10 + 441437.0	8	161.40	+0.11	16475 ± 22	47 ± 66	19 ± 88	2.10 ± 0.41	0.04 ± 0.12	0.89	235	8.70	–
115	J045254.20 + 442339.7	8	161.33	+0.26	9942 ± 11	207 ± 33	190 ± 44	1.30 ± 0.41	0.03 ± 0.31	0.24	142	8.19	–
116	J045304.54 + 442608.1	8	161.31	+0.31	16473 ± 03	106 ± 09	58 ± 12	2.60 ± 0.42	0.50 ± 0.06	6.63	235	9.81	–
117	J045307.38 + 441620.6	8	161.45	+0.21	16477 ± 47	154 ± 141	27. ± 188	1.80 ± 0.41	0.03 ± 0.13	0.68	235	8.66	–
118	J045420.60 + 454724.2	5	160.41	+1.34	10815 ± 01	151 ± 03	119 ± 04	2.70 ± 0.42	0.51 ± 0.09	4.75	155	9.46	–
119	J045449.71 + 462050.9	2	160.03	+1.76	9457 ± 10	84 ± 30	42 ± 40	4.20 ± 0.45	0.07 ± 0.17	1.16	135	8.50	–
120	J045557.87 + 464609.4	2	159.82	+2.17	7078 ± 01	138 ± 03	120 ± 04	2.40 ± 0.42	0.35 ± 0.10	3.22	101	8.92	–
121	J045605.93 + 435110.4	8	162.11	+0.37	13750 ± 28	74 ± 84	17. ± 112	1.90 ± 0.41	0.01 ± 0.27	0.14	196	7.71	–
122	J045626.94 + 455542.0	2	160.53	+1.71	6886 ± 15	985 ± 45	526 ± 60	5.00 ± 0.47	0.46 ± 0.27	2.05	98	9.02	–
123	J045629.65 + 440244.9	8	162.00	+0.54	3880 ± 02	80 ± 06	53 ± 08	2.40 ± 0.42	0.29 ± 0.07	4.00	55	8.32	–
124	J045742.64 + 461101.0	2	160.47	+2.04	8342 ± 07	314 ± 21	305 ± 28	1.50 ± 0.41	0.05 ± 0.36	0.27	119	8.18	–
125	J045755.76 + 461625.6	2	160.42	+2.13	6220 ± 02	29 ± 06	22 ± 08	1.00 ± 0.40	0.10 ± 0.05	2.26	89	8.29	–
126	J045808.00 + 460927.9	1	160.53	+2.08	8298 ± 24	123 ± 72	101 ± 96	1.80 ± 0.41	0.02 ± 0.31	0.25	119	7.91	–
127	J045939.30 + 442408.5	7	162.08	+1.21	15691 ± 33	1785 ± 99	750 ± 132	2.30 ± 0.42	1.93 ± 0.11	7.16	224	10.36	–
128	J050137.02 + 450141.9	4	161.80	+1.87	14851 ± 01	39 ± 03	32 ± 04	2.90 ± 0.43	0.17 ± 0.08	3.04	212	9.25	–
129	J050222.78 + 441926.6	7	162.44	+1.54	15232 ± 13	267 ± 39	50 ± 52	2.10 ± 0.41	0.14 ± 0.10	2.07	218	9.21	–
130	J050231.21 + 450323.1	4	161.87	+2.01	14855 ± 10	106 ± 30	11 ± 40	2.20 ± 0.41	0.09 ± 0.06	2.72	212	8.97	–
131	J050332.64 + 451201.6	4	161.87	+2.24	14849 ± 04	166 ± 12	98 ± 16	2.60 ± 0.42	0.32 ± 0.10	3.30	212	9.53	–

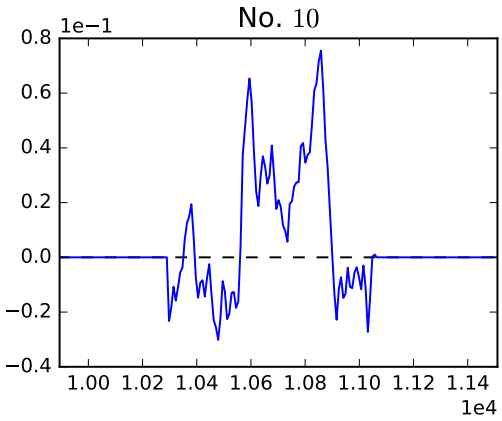
Appendix C

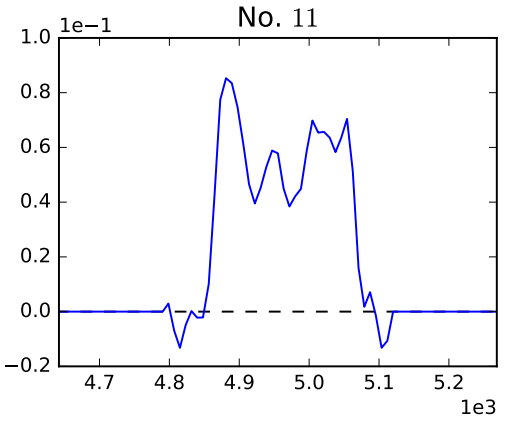
Global profiles of SoFiA detections

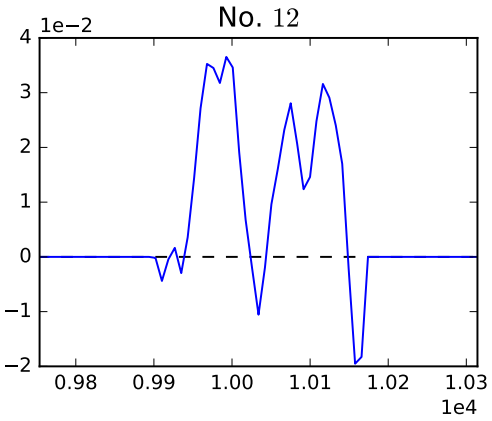
Presented here are global profiles of the 131 SoFiA detections. The names of the figures (e.g. No. 1) correspond to the numbers in the first column of B. 1. All the profiles of detections with both Visual and SA crossmatches were plotted in blue, whereas red show those with no crossmatches from the Testbed sample. It must be noted that the baseline has been subtracted already on all the spectra. On the x-axis is the heliocentric velocity in $\mathrm{km\,s^{-1}}$ and the y-axis is the sum of flux values across the area covered by the source and has not been multiplied by the beam area, hence the units are $\mathrm{Jy\ beam^{-1}}$.

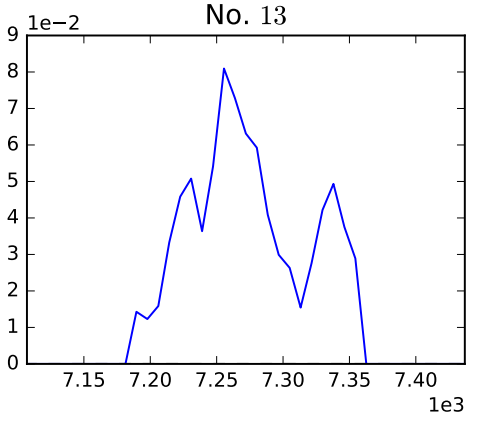

No. 1
1e−2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
4.98 5.00 5.02 5.04 5.06 5.08 5.10 5.12
1e3
No. 2
1e−2
4
3
2
1
0
−1
−2
−3
1.02 1.04 1.06 1.08 1.10 1.12
1e4
No. 3
1e−2
6
5
4
3
2
1
0
−1
6.65 6.70 6.75 6.80 6.85 6.90
1e3
No. 4
1e−1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
6.1 6.2 6.3 6.4
1e3
No. 5
1e−1
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.94 0.95 0.96 0.97 0.98 0.99 1.00
1e4
No. 6
1e−2
4
3
2
1
0
−1
0.9800.9850.9900.9951.0001.0051.0101.015
1e4
No. 7
1e−2
6
5
4
3
2
1
0
−1
5.4 5.5 5.6 5.7 5.8
1e3
No. 8
1e−2
5
4
3
2
1
0
−1
−2
5.50 5.55 5.60 5.65 5.70 5.75 5.80 5.85
1e3

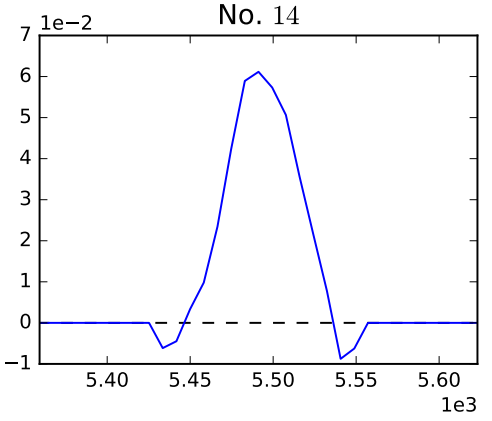

No. 9
1e−2
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
7.30 7.35 7.40 7.45 7.50 7.55 7.60 7.65
1e3

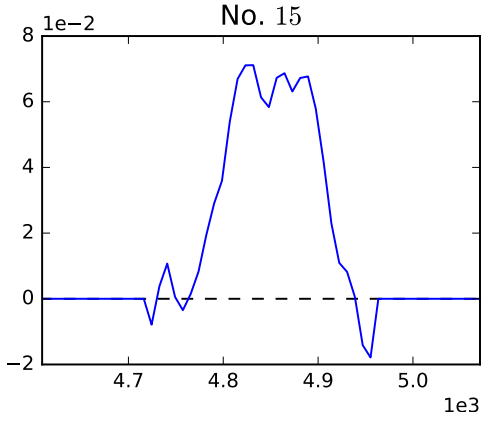

No. 10
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
1.00 1.02 1.04 1.06 1.08 1.10 1.12 1.14
1e4

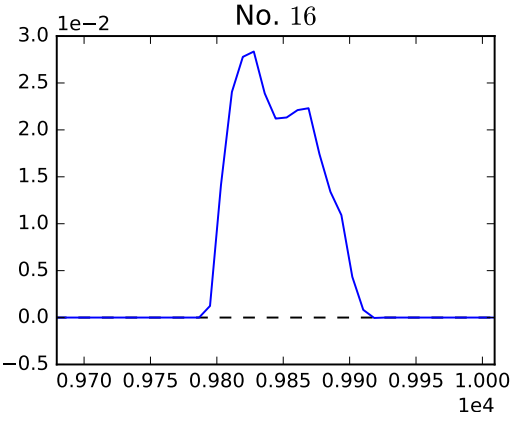

No. 11
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
4.7 4.8 4.9 5.0 5.1 5.2
1e3

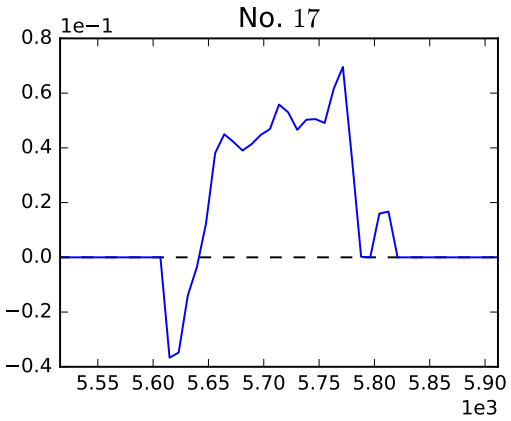

No. 12
1e−2
4
3
2
1
0
−1
−2
0.98 0.99 1.00 1.01 1.02 1.03
1e4

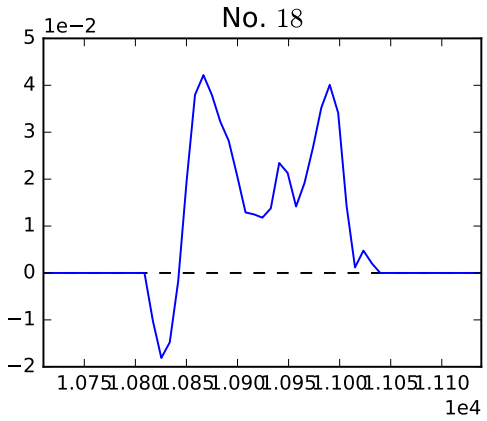

No. 13
1e−2
9
8
7
6
5
4
3
2
1
0
7.15 7.20 7.25 7.30 7.35 7.40
1e3

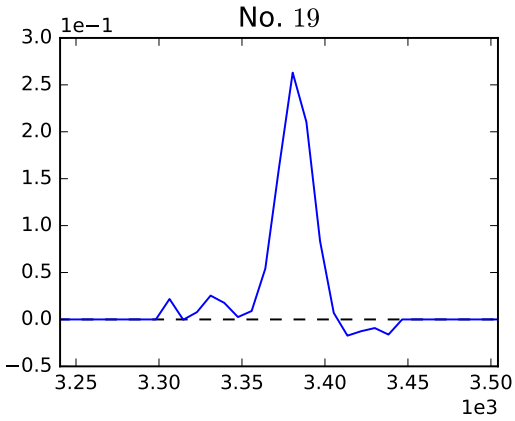

No. 14
1e−2
7
6
5
4
3
2
1
0
−1
5.40 5.45 5.50 5.55 5.60
1e3

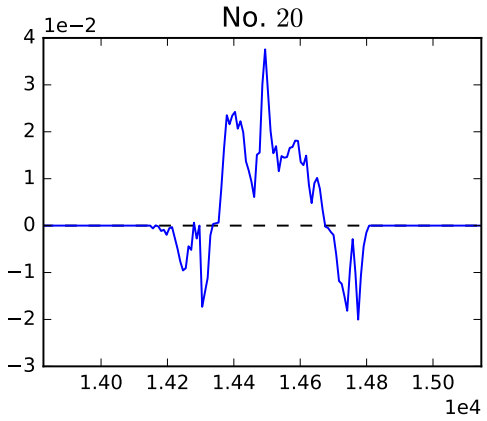

No. 15
1e−2
8
6
4
2
0
−2
4.7 4.8 4.9 5.0
1e3

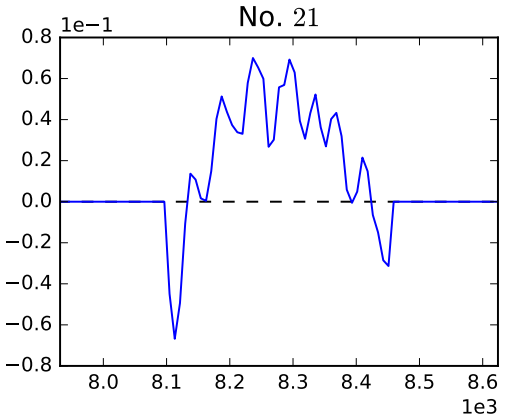

No. 16
1e−2
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0.970 0.975 0.980 0.985 0.990 0.995 1.000
1e4

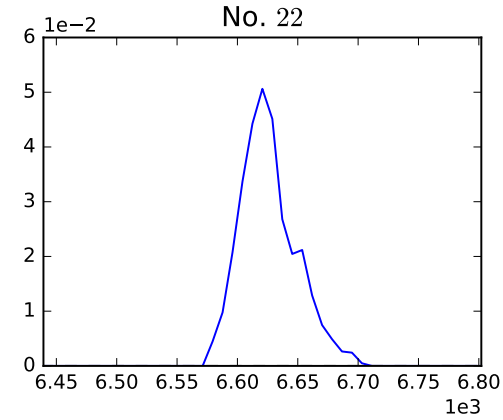

No. 17
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
5.55 5.60 5.65 5.70 5.75 5.80 5.85 5.90
1e3

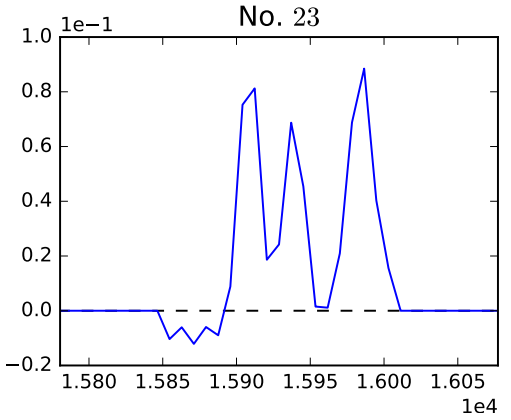

No. 18
1e−2
5
4
3
2
1
0
−1
−2
1.075 1.080 1.085 1.090 1.095 1.100 1.105 1.110
1e4

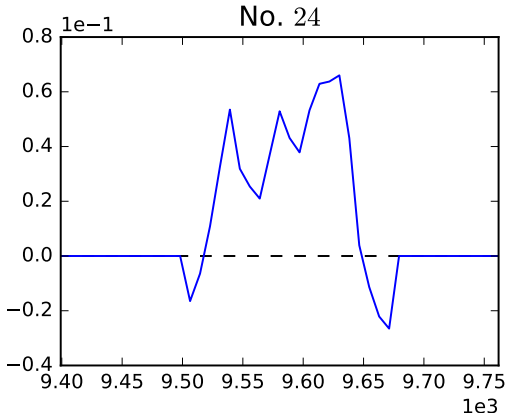

No. 19
1e−1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
3.25 3.30 3.35 3.40 3.45 3.50
1e3

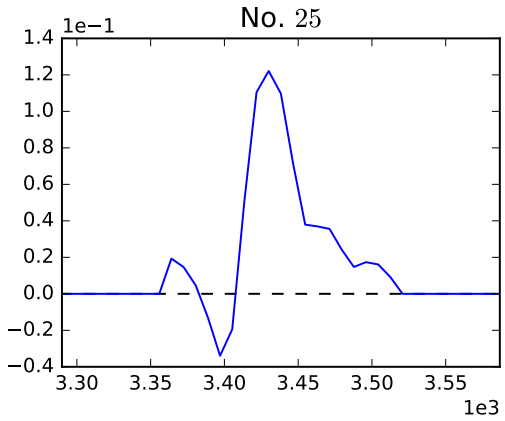

No. 20
1e−2
4
3
2
1
0
−1
−2
−3
1.40 1.42 1.44 1.46 1.48 1.50
1e4

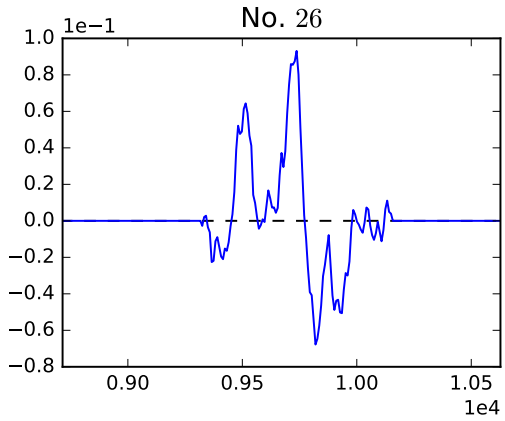

No. 21
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
8.0 8.1 8.2 8.3 8.4 8.5 8.6
1e3

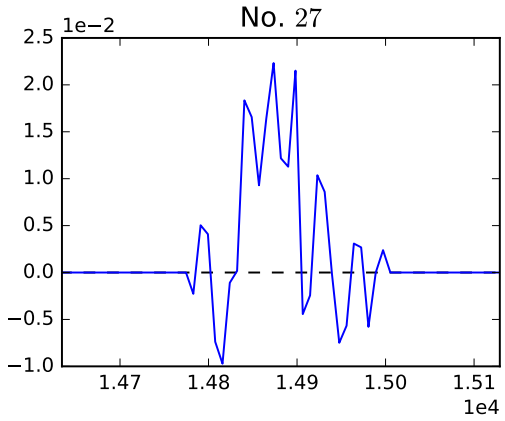

No. 22
1e−2
6
5
4
3
2
1
0
6.45 6.50 6.55 6.60 6.65 6.70 6.75 6.80
1e3

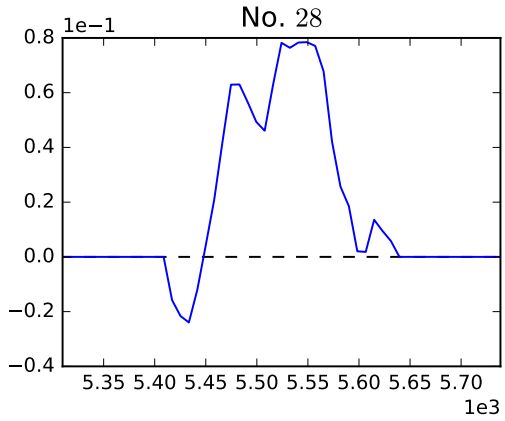

No. 23
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
1.580 1.585 1.590 1.595 1.600 1.605
1e4

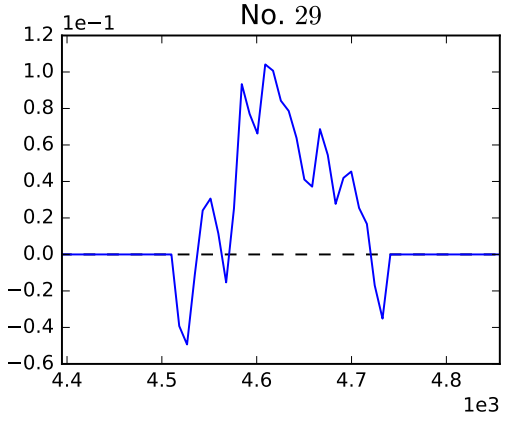

No. 24
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
9.40 9.45 9.50 9.55 9.60 9.65 9.70 9.75
1e3

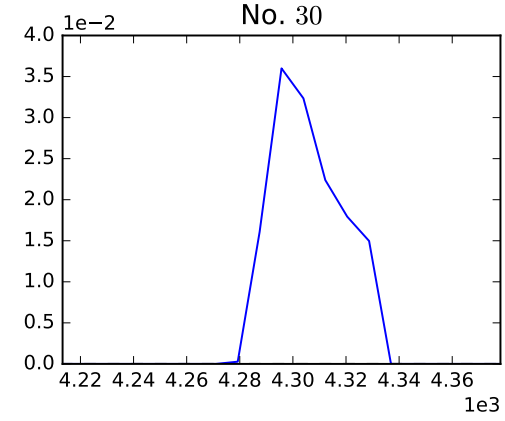

No. 25
1e−1
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
3.30 3.35 3.40 3.45 3.50 3.55
1e3

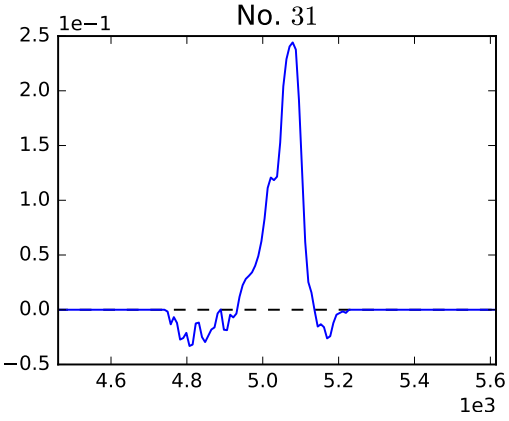

No. 26
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.90 0.95 1.00 1.05
1e4

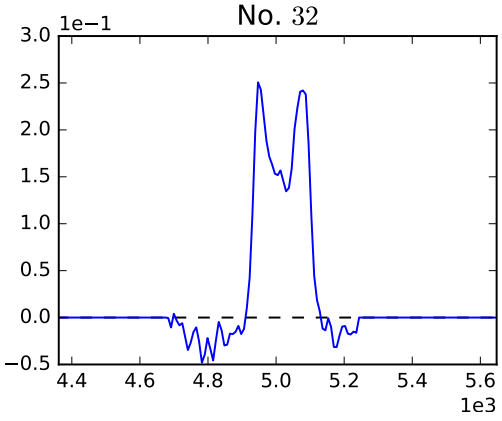

No. 27
1e−2
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
1.47 1.48 1.49 1.50 1.51
1e4

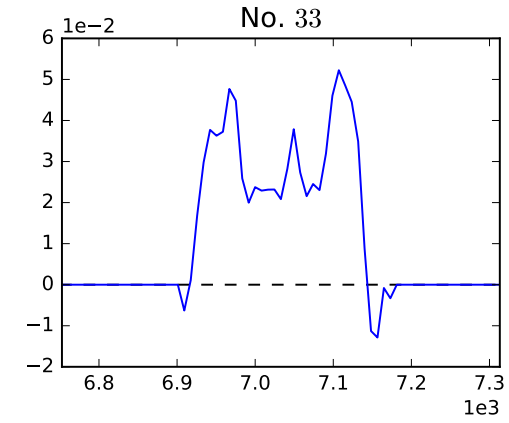

No. 28
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
5.35 5.40 5.45 5.50 5.55 5.60 5.65 5.70
1e3

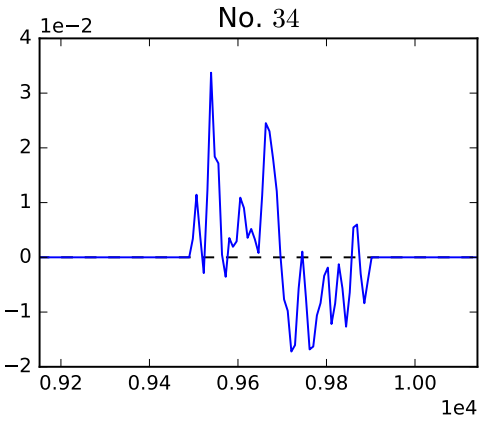

No. 29
1e−1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
4.4 4.5 4.6 4.7 4.8
1e3

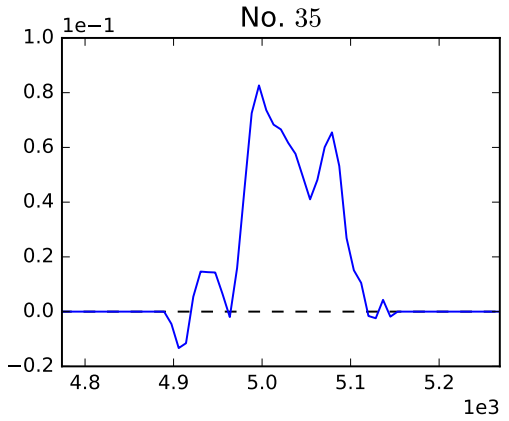

No. 30
1e−2
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
4.22 4.24 4.26 4.28 4.30 4.32 4.34 4.36
1e3

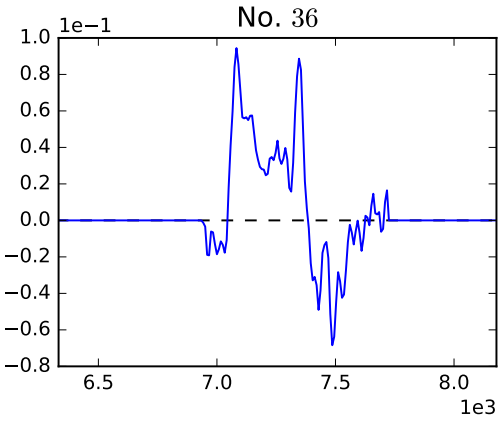

No. 31
1e−1
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
4.6 4.8 5.0 5.2 5.4 5.6
1e3

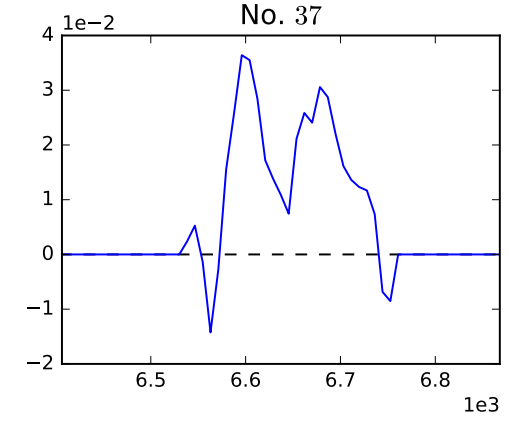

No. 32
1e−1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
4.4 4.6 4.8 5.0 5.2 5.4 5.6
1e3

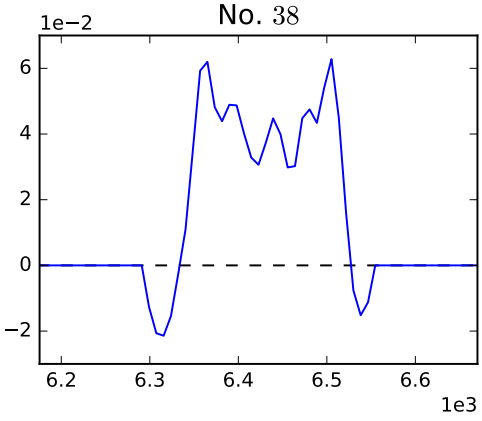

No. 33
1e−2
6
5
4
3
2
1
0
−1
−2
6.8
6.9
7.0
7.1
7.2
7.3
1e3

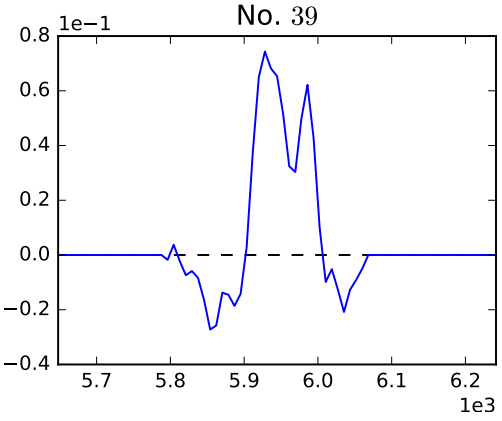

No. 34
1e−2
4
3
2
1
0
−1
−2
0.92
0.94
0.96
0.98
1.00
1e4

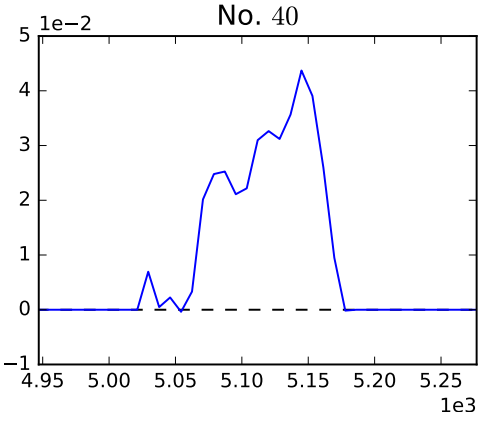

No. 35
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
4.8
4.9
5.0
5.1
5.2
1e3

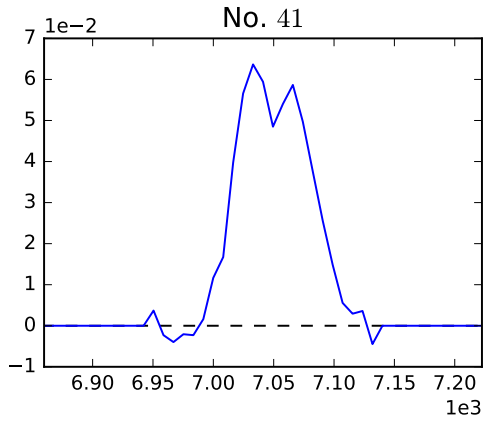

No. 36
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
6.5
7.0
7.5
8.0
1e3

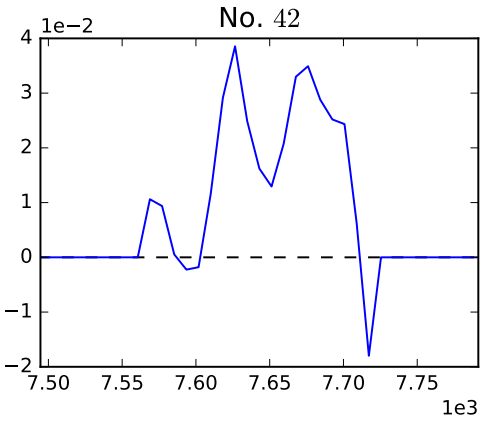

No. 37
1e−2
4
3
2
1
0
−1
−2
6.5
6.6
6.7
6.8
1e3

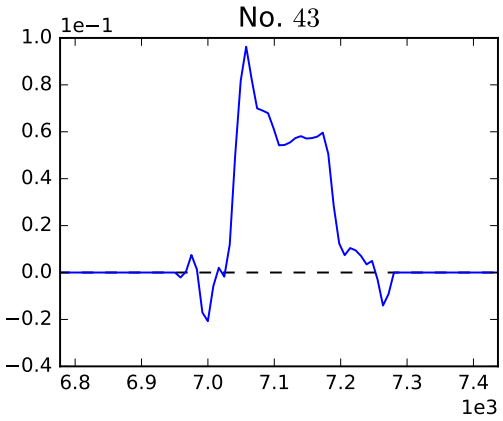

No. 38
1e−2
6
4
2
0
−2
6.2
6.3
6.4
6.5
6.6
1e3

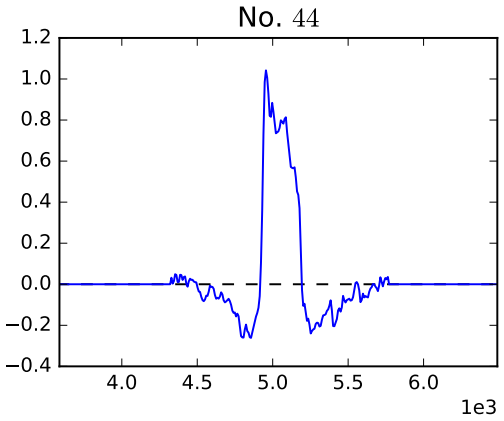

No. 39
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
5.7
5.8
5.9
6.0
6.1
6.2
1e3

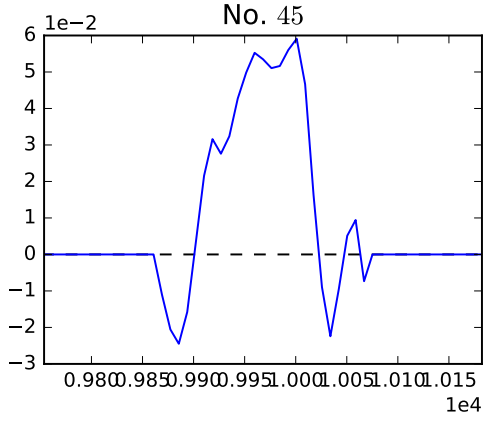

No. 40
1e−2
5
4
3
2
1
0
−1
4.95
5.00
5.05
5.10
5.15
5.20
5.25
1e3

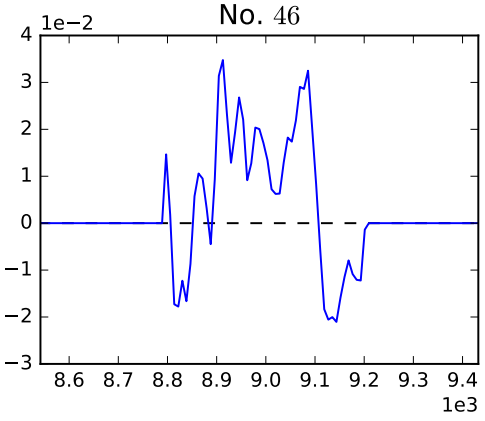

No. 41
1e−2
7
6
5
4
3
2
1
0
−1
6.90 6.95 7.00 7.05 7.10 7.15 7.20
1e3

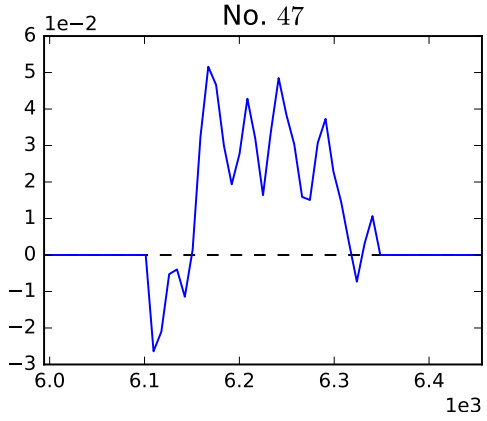

No. 42
1e−2
4
3
2
1
0
−1
−2
7.50 7.55 7.60 7.65 7.70 7.75
1e3

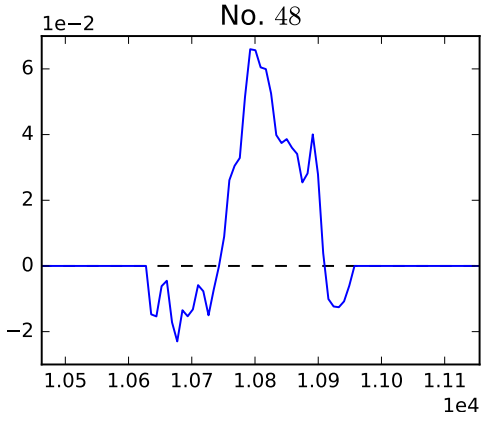

No. 43
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
6.8 6.9 7.0 7.1 7.2 7.3 7.4
1e3

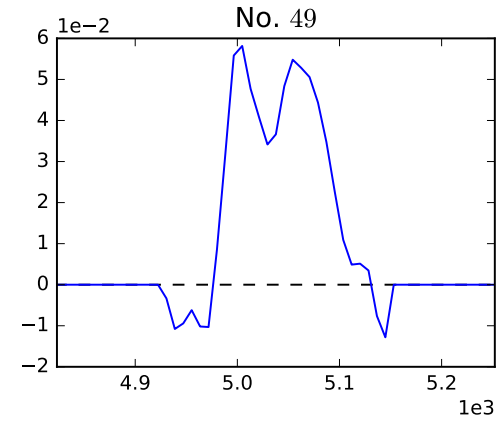

No. 44
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
4.0 4.5 5.0 5.5 6.0
1e3

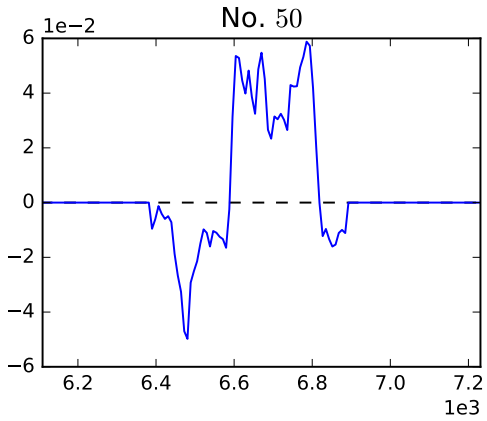

No. 45
1e−2
6
5
4
3
2
1
0
−1
−2
−3
0.980 0.985 0.990 0.995 1.000 1.005 1.010 1.015
1e4

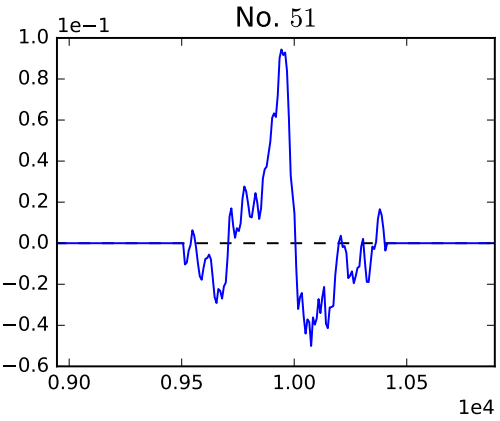

No. 46
1e−2
4
3
2
1
0
−1
−2
−3
8.6 8.7 8.8 8.9 9.0 9.1 9.2 9.3 9.4
1e3

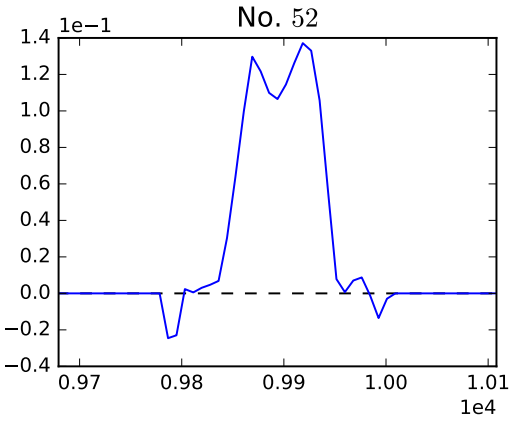

No. 47
1e−2
6
5
4
3
2
1
0
−1
−2
−3
6.0 6.1 6.2 6.3 6.4
1e3

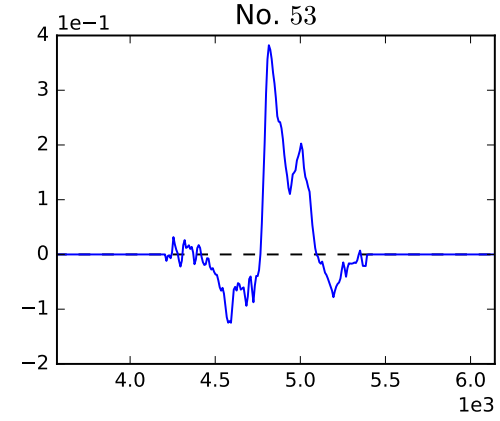

No. 48
1e−2
6
4
2
0
−2
1.05 1.06 1.07 1.08 1.09 1.10 1.11
1e4

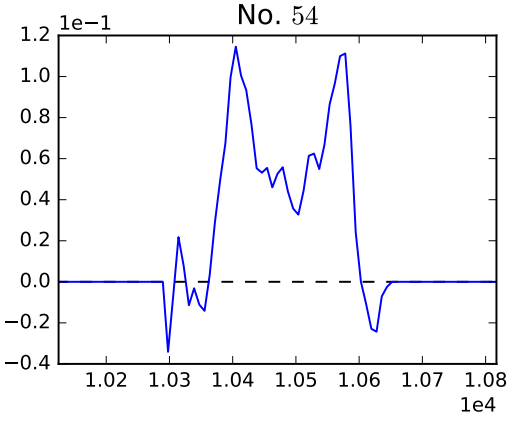

No. 49
1e−2
6
5
4
3
2
1
0
−1
−2
4.9
5.0
5.1
5.2
1e3

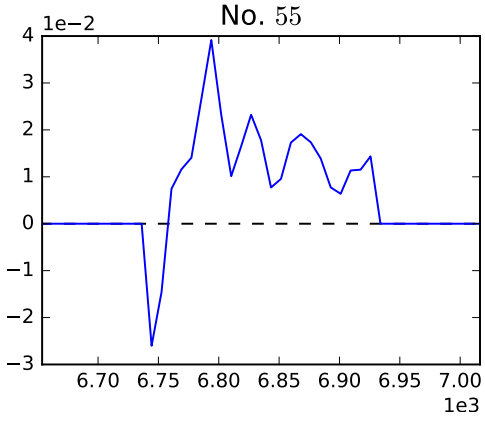

No. 50
1e−2
6
4
2
0
−2
−4
−6
6.2
6.4
6.6
6.8
7.0
7.2
1e3

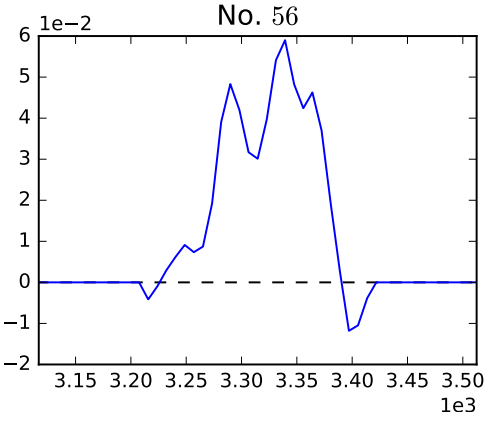

No. 51
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.90
0.95
1.00
1.05
1e4

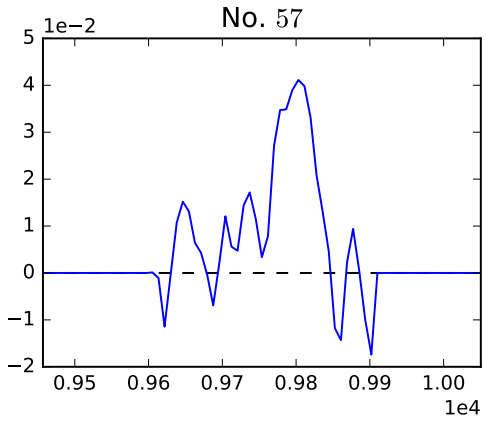

No. 52
1e−1
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.97
0.98
0.99
1.00
1.01
1e4

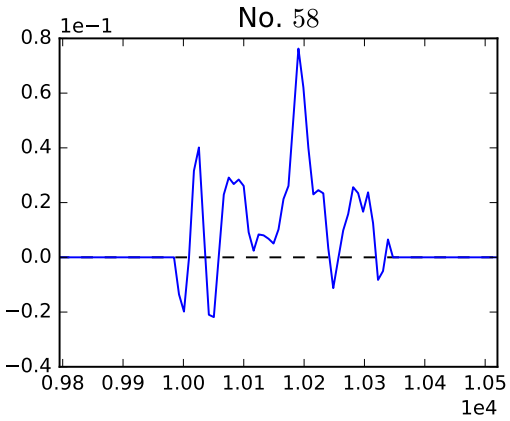

No. 53
1e−1
4
3
2
1
0
−1
−2
4.0
4.5
5.0
5.5
6.0
1e3

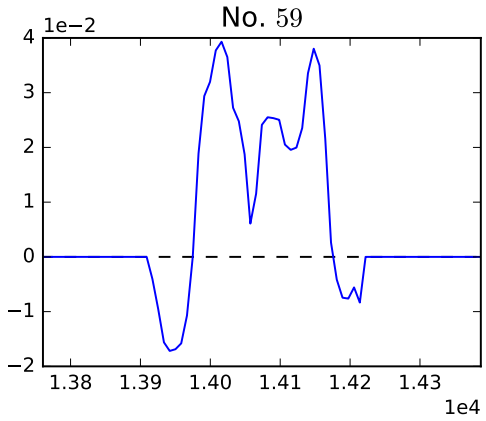

No. 54
1e−1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
1.02
1.03
1.04
1.05
1.06
1.07
1.08
1e4

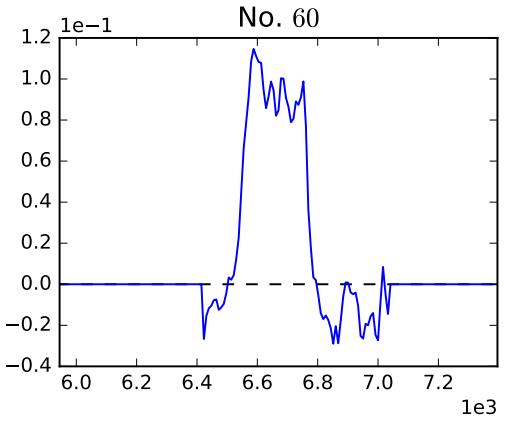

No. 55
1e−2
4
3
2
1
0
−1
−2
−3
6.70
6.75
6.80
6.85
6.90
6.95
7.00
1e3

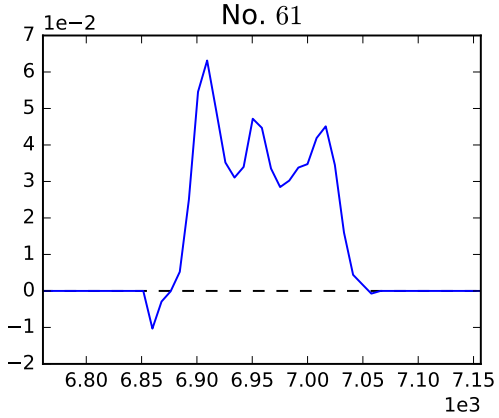

No. 56
1e−2
6
5
4
3
2
1
0
−1
−2
3.15
3.20
3.25
3.30
3.35
3.40
3.45
3.50
1e3

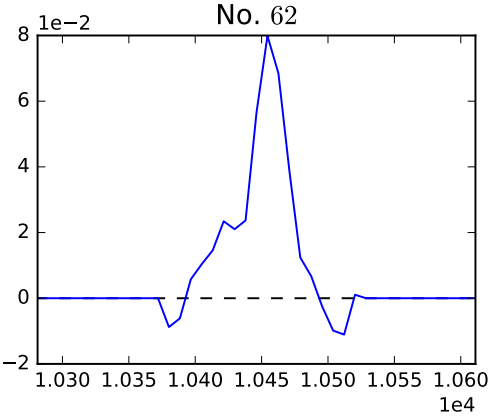

No. 57
1e−2
5
4
3
2
1
0
−1
−2
0.95 0.96 0.97 0.98 0.99 1.00
1e4

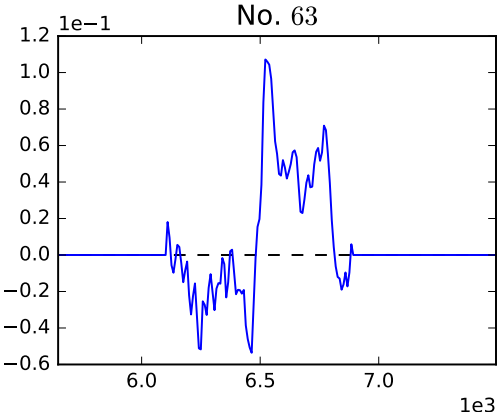

No. 58
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.98 0.99 1.00 1.01 1.02 1.03 1.04 1.05
1e4

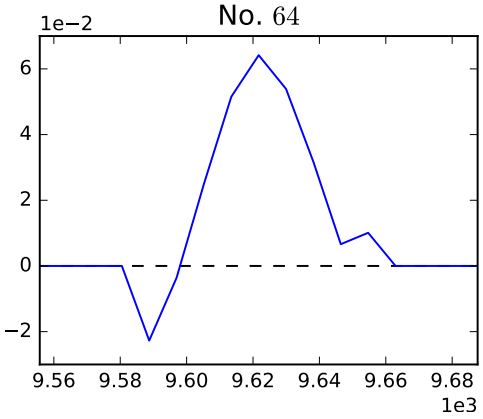

No. 59
1e−2
4
3
2
1
0
−1
−2
1.38 1.39 1.40 1.41 1.42 1.43
1e4

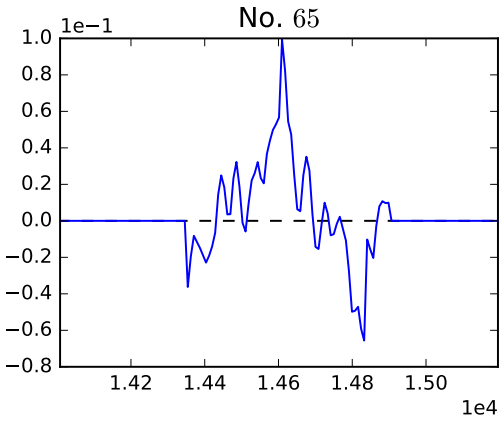

No. 60
1e−1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
6.0 6.2 6.4 6.6 6.8 7.0 7.2
1e3

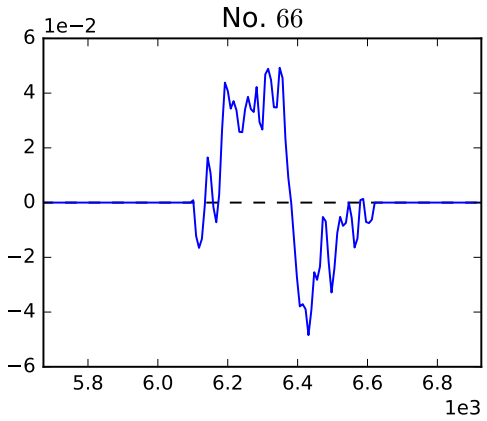

No. 61
1e−2
7
6
5
4
3
2
1
0
−1
−2
6.80 6.85 6.90 6.95 7.00 7.05 7.10 7.15
1e3

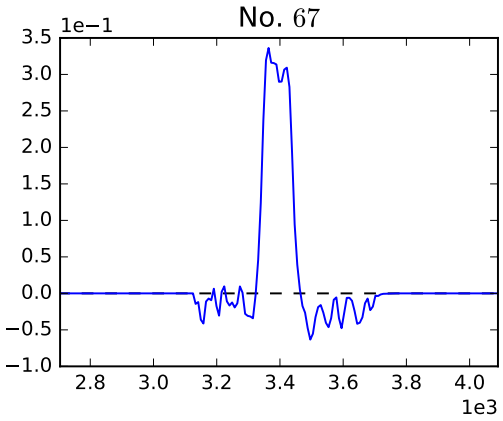

No. 62
1e−2
8
6
4
2
0
−2
1.030 1.035 1.040 1.045 1.050 1.055 1.060
1e4

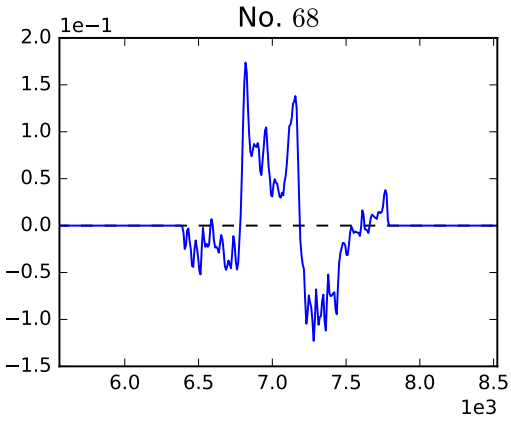

No. 63
1e−1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
6.0 6.5 7.0
1e3

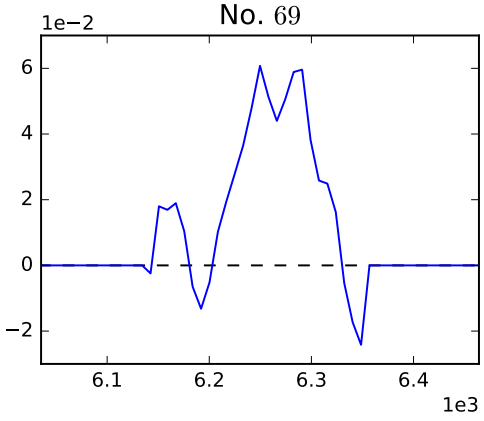

No. 64
1e−2
6
4
2
0
−2
9.56 9.58 9.60 9.62 9.64 9.66 9.68
1e3

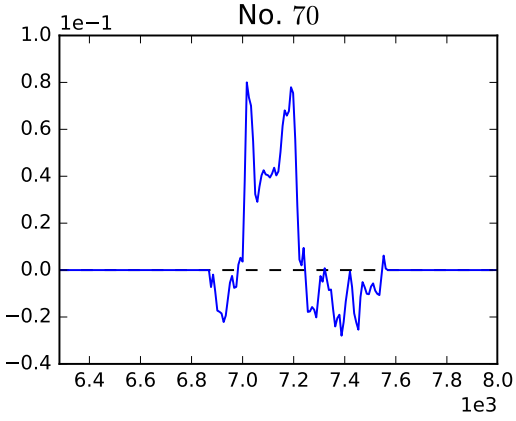

No. 65
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
1.42
1.44
1.46
1.48
1.50
1e4

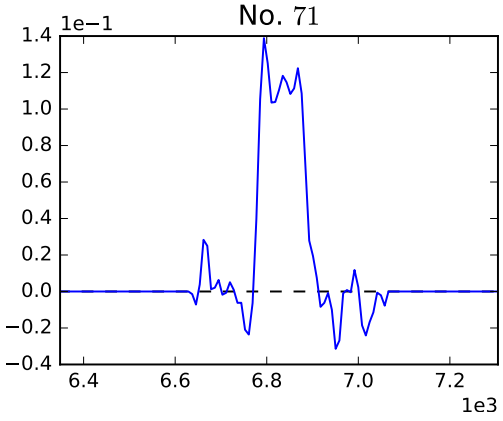

No. 66
1e−2
6
4
2
0
−2
−4
−6
5.8
6.0
6.2
6.4
6.6
6.8
1e3

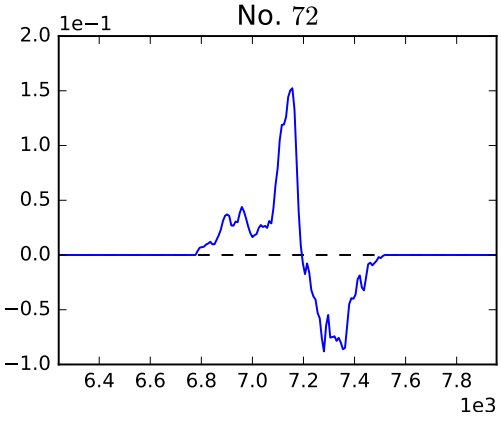

No. 67
1e−1
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
2.8
3.0
3.2
3.4
3.6
3.8
4.0
1e3

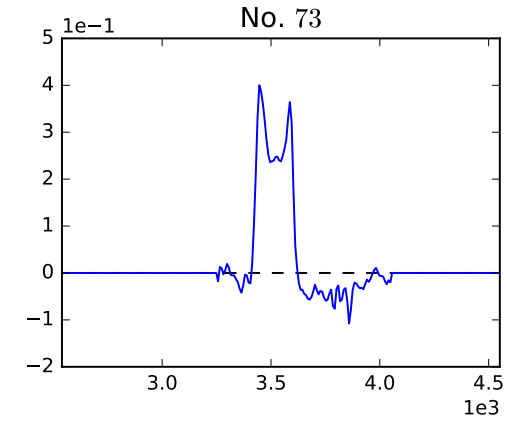

No. 68
1e−1
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
6.0
6.5
7.0
7.5
8.0
8.5
1e3

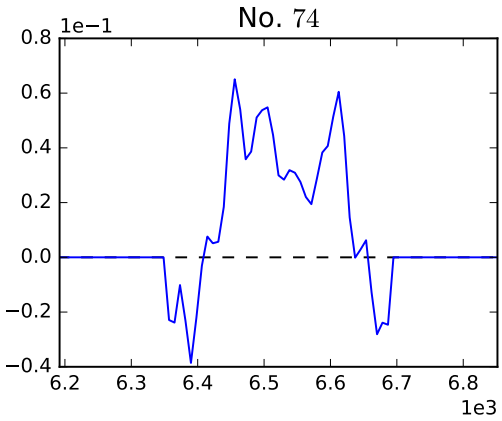

No. 69
1e−2
6
4
2
0
−2
6.1
6.2
6.3
6.4
1e3

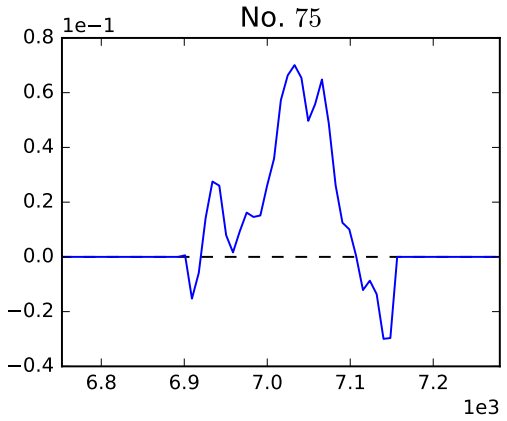

No. 70
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
6.4
6.6
6.8
7.0
7.2
7.4
7.6
7.8
8.0
1e3

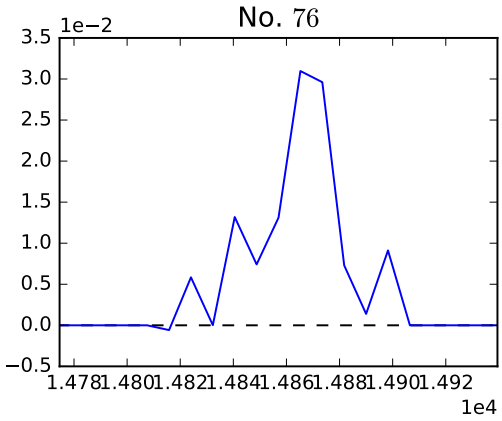

No. 71
1e−1
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
6.4
6.6
6.8
7.0
7.2
1e3

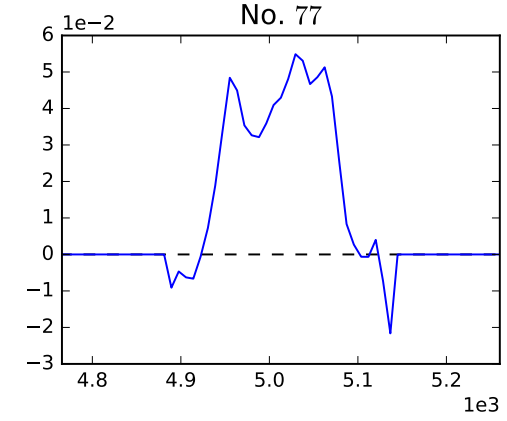

No. 72
1e−1
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
6.4
6.6
6.8
7.0
7.2
7.4
7.6
7.8
1e3

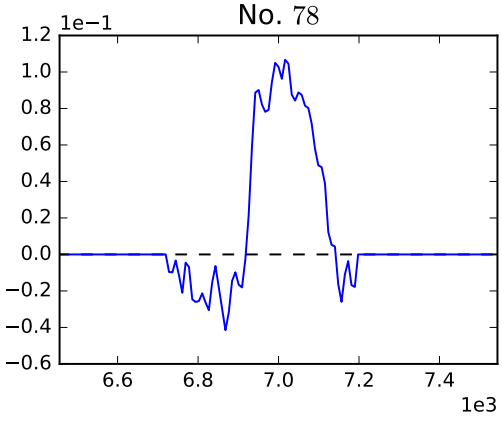

No. 73
1e−1
5
4
3
2
1
0
−1
−2
3.0
3.5
4.0
4.5
1e3

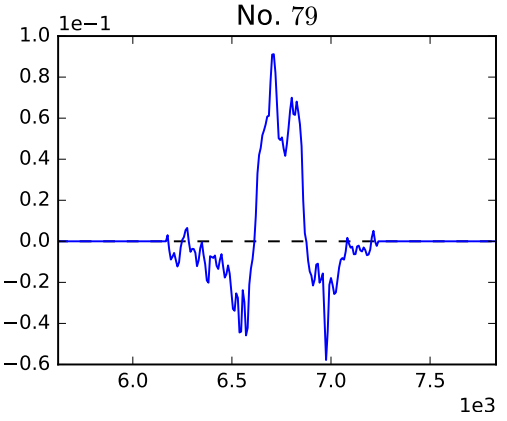

No. 74
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
6.2
6.3
6.4
6.5
6.6
6.7
6.8
1e3

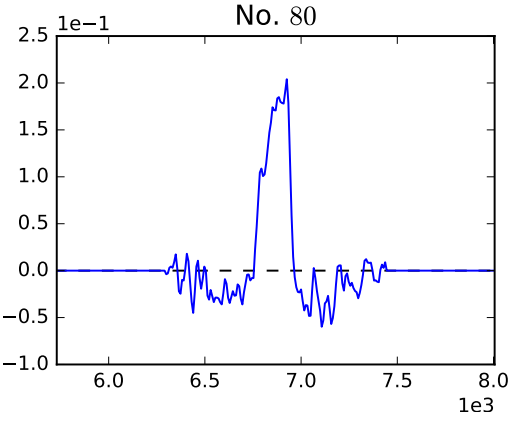

No. 75
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
6.8
6.9
7.0
7.1
7.2
1e3

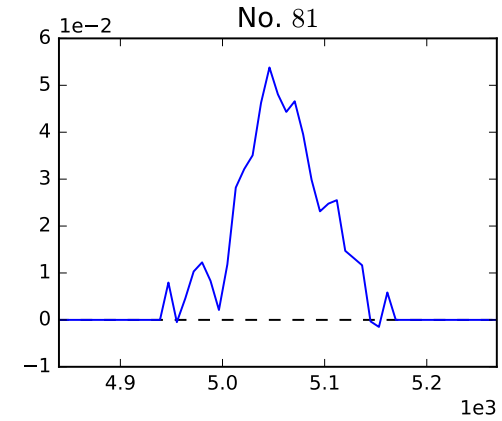

No. 76
1e−2
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
1.478
1.480
1.482
1.484
1.486
1.488
1.490
1.492
1e4

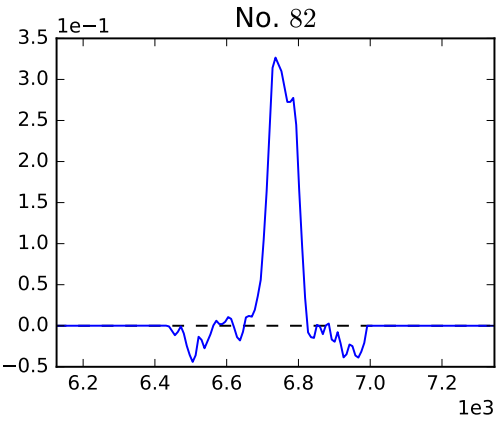

No. 77
1e−2
6
5
4
3
2
1
0
−1
−2
−3
4.8
4.9
5.0
5.1
5.2
1e3

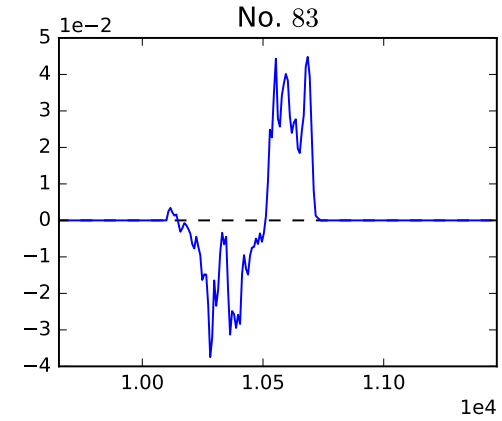

No. 78
1e−1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
6.6
6.8
7.0
7.2
7.4
1e3

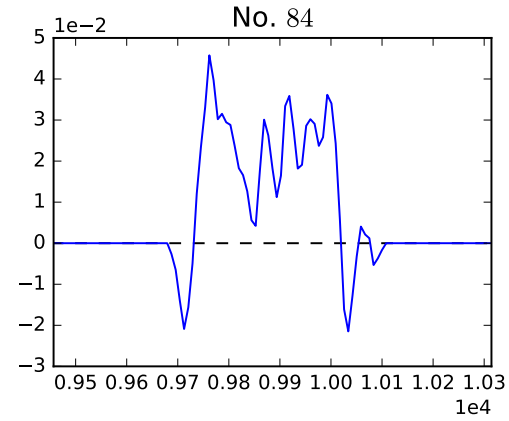

No. 79
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
6.0
6.5
7.0
7.5
1e3

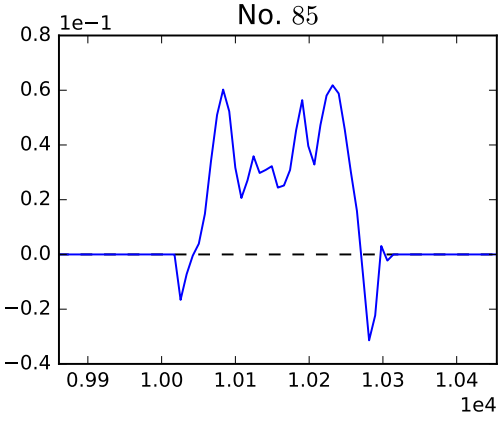

No. 80
1e−1
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
6.0
6.5
7.0
7.5
8.0
1e3

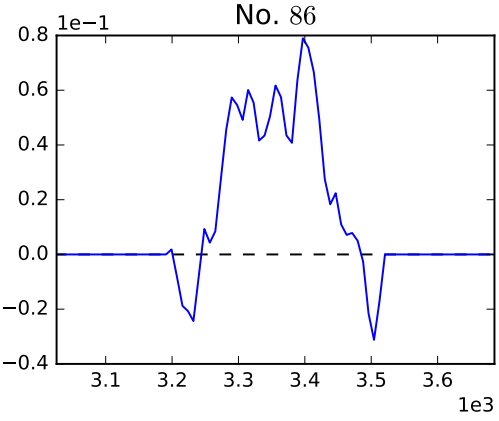

No. 81
1e−2
6
5
4
3
2
1
0
−1
4.9
5.0
5.1
5.2
1e3

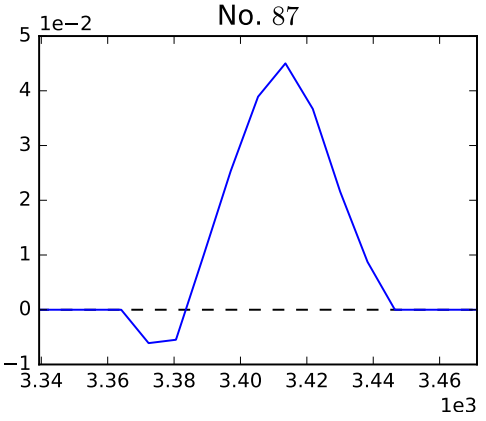

No. 82
1e−1
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
6.2
6.4
6.6
6.8
7.0
7.2
1e3

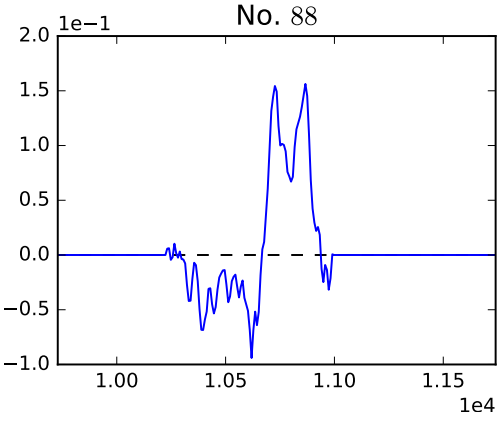

No. 83
1e−2
5
4
3
2
1
0
−1
−2
−3
−4
1.00
1.05
1.10
1e4

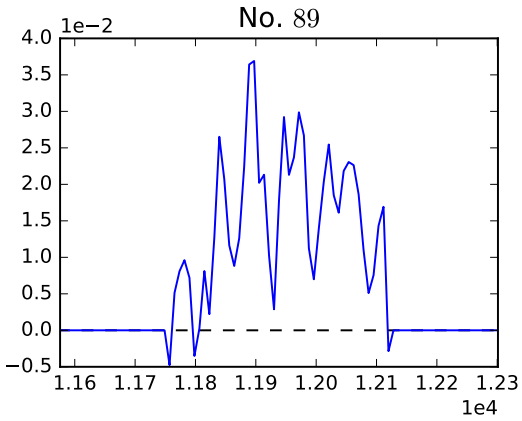

No. 84
1e−2
5
4
3
2
1
0
−1
−2
−3
0.95 0.96 0.97 0.98 0.99 1.00 1.01 1.02 1.03
1e4

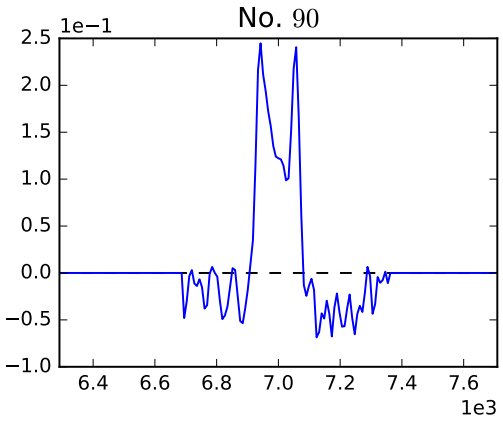

No. 85
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.99
1.00
1.01
1.02
1.03
1.04
1e4

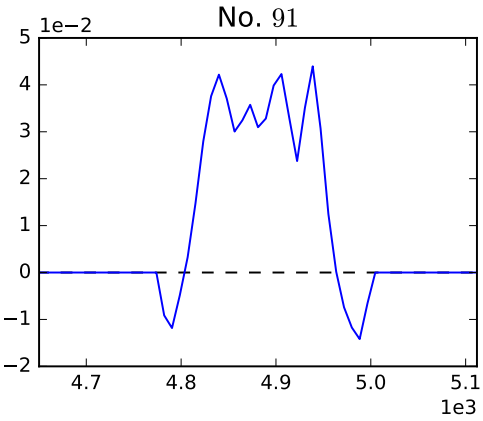

No. 86
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
3.1
3.2
3.3
3.4
3.5
3.6
1e3

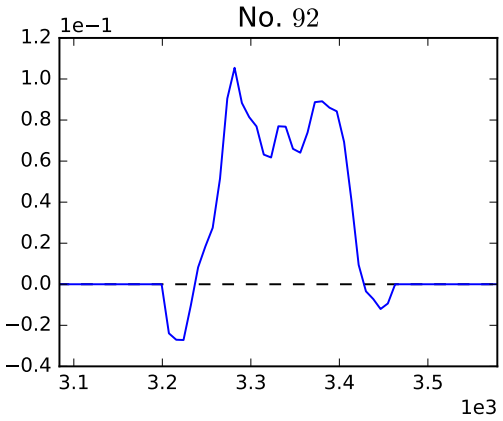

No. 87
1e−2
5
4
3
2
1
0
−1
3.34
3.36
3.38
3.40
3.42
3.44
3.46
1e3

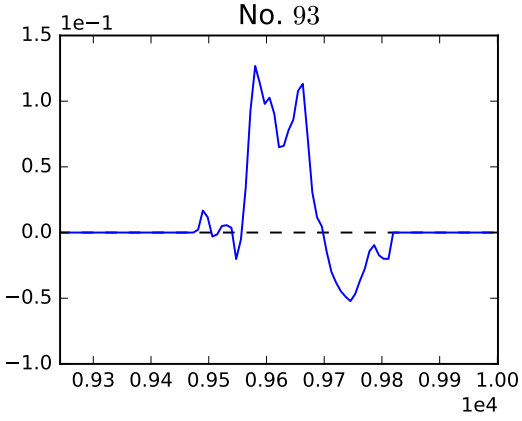

No. 88
1e−1
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
1.00
1.05
1.10
1.15
1e4

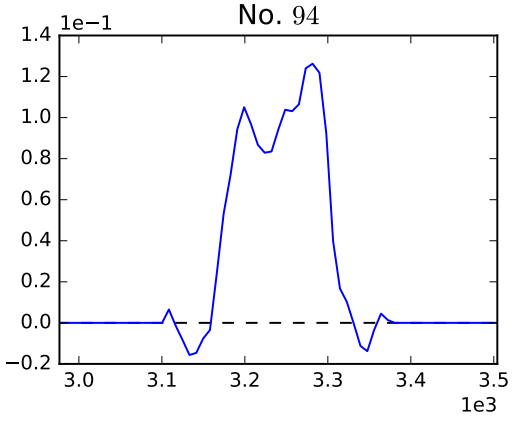

No. 89
1e−2
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
1.16 1.17 1.18 1.19 1.20 1.21 1.22 1.23
1e4

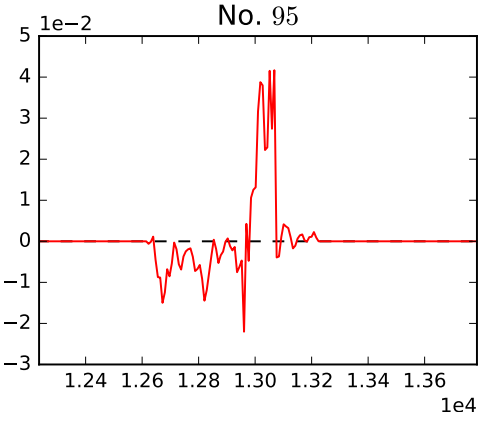

No. 90
1e−1
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
6.4 6.6 6.8 7.0 7.2 7.4 7.6
1e3

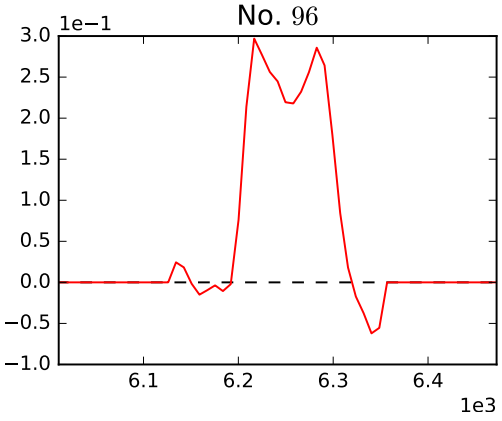

No. 91
1e−2
5
4
3
2
1
0
−1
−2
4.7 4.8 4.9 5.0 5.1
1e3

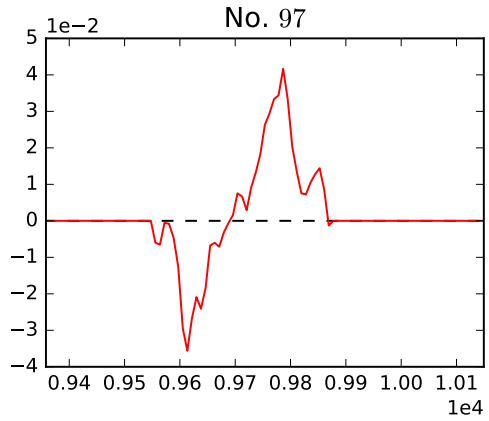

No. 92
1e−1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
3.1 3.2 3.3 3.4 3.5
1e3

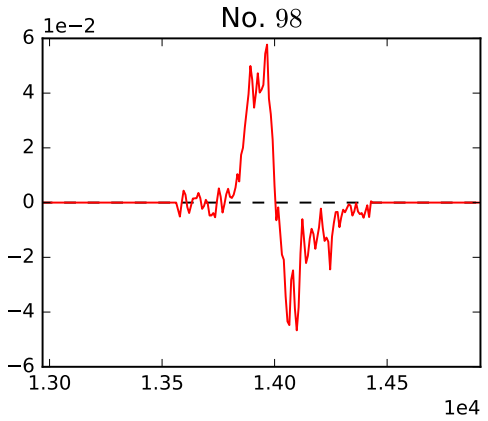

No. 93
1e−1
1.5
1.0
0.5
0.0
−0.5
−1.0
0.93 0.94 0.95 0.96 0.97 0.98 0.99 1.00
1e4

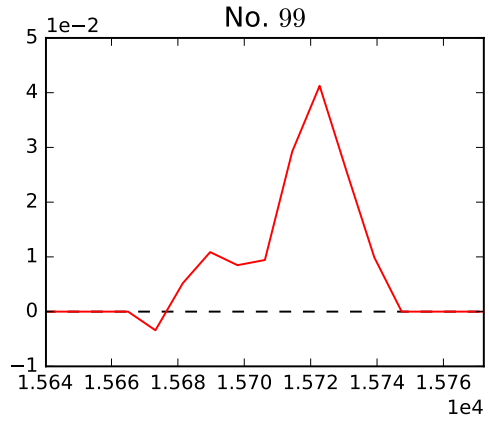

No. 94
1e−1
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
3.0 3.1 3.2 3.3 3.4 3.5
1e3

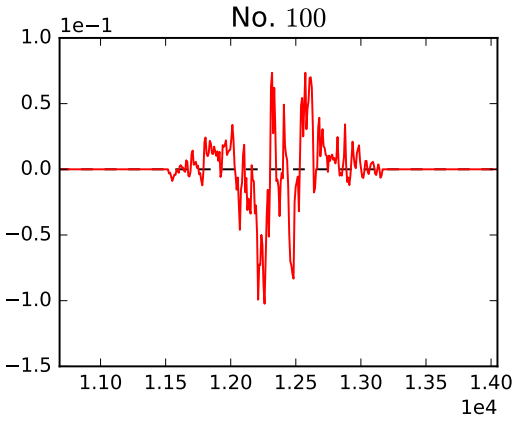

No. 95
1e−2
5
4
3
2
1
0
−1
−2
−3
1.24 1.26 1.28 1.30 1.32 1.34 1.36
1e4

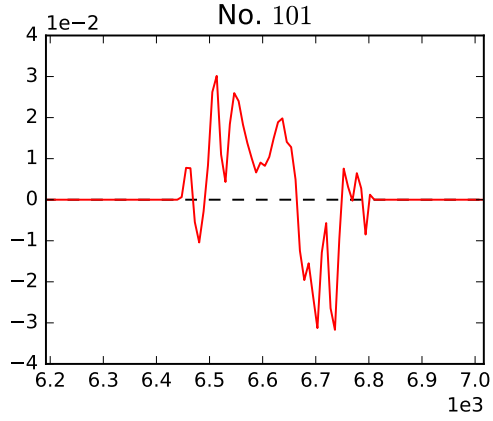

No. 96
1e−1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
6.1 6.2 6.3 6.4
1e3

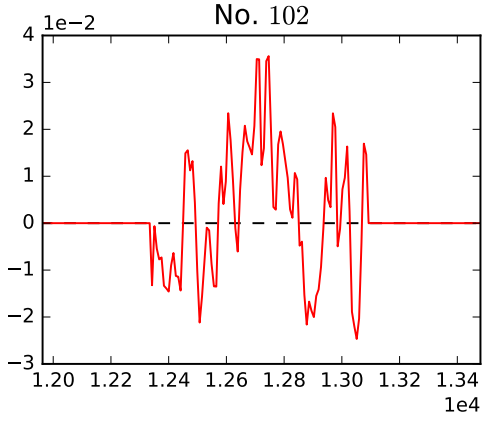

No. 97
1e−2
5
4
3
2
1
0
−1
−2
−3
−4
0.94 0.95 0.96 0.97 0.98 0.99 1.00 1.01
1e4

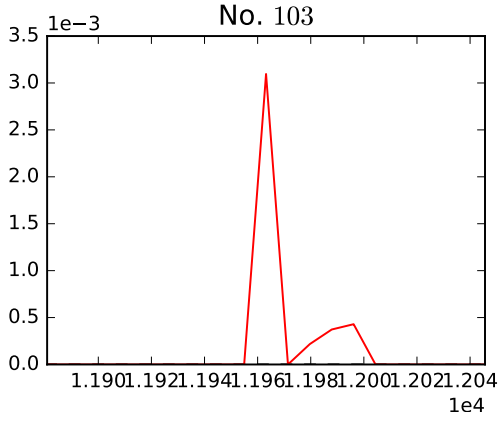

No. 98
1e−2
6
4
2
0
−2
−4
−6
1.30 1.35 1.40 1.45
1e4

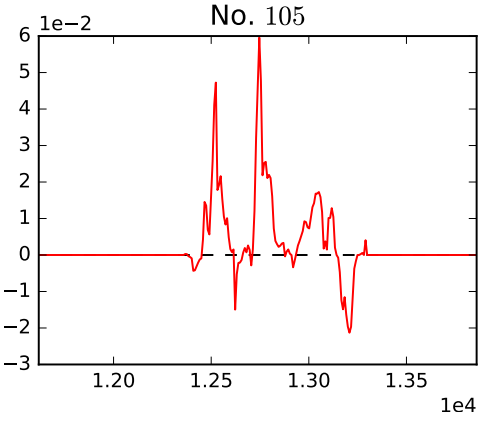

No. 99
1e−2
5
4
3
2
1
0
−1
1.564 1.566 1.568 1.570 1.572 1.574 1.576
1e4

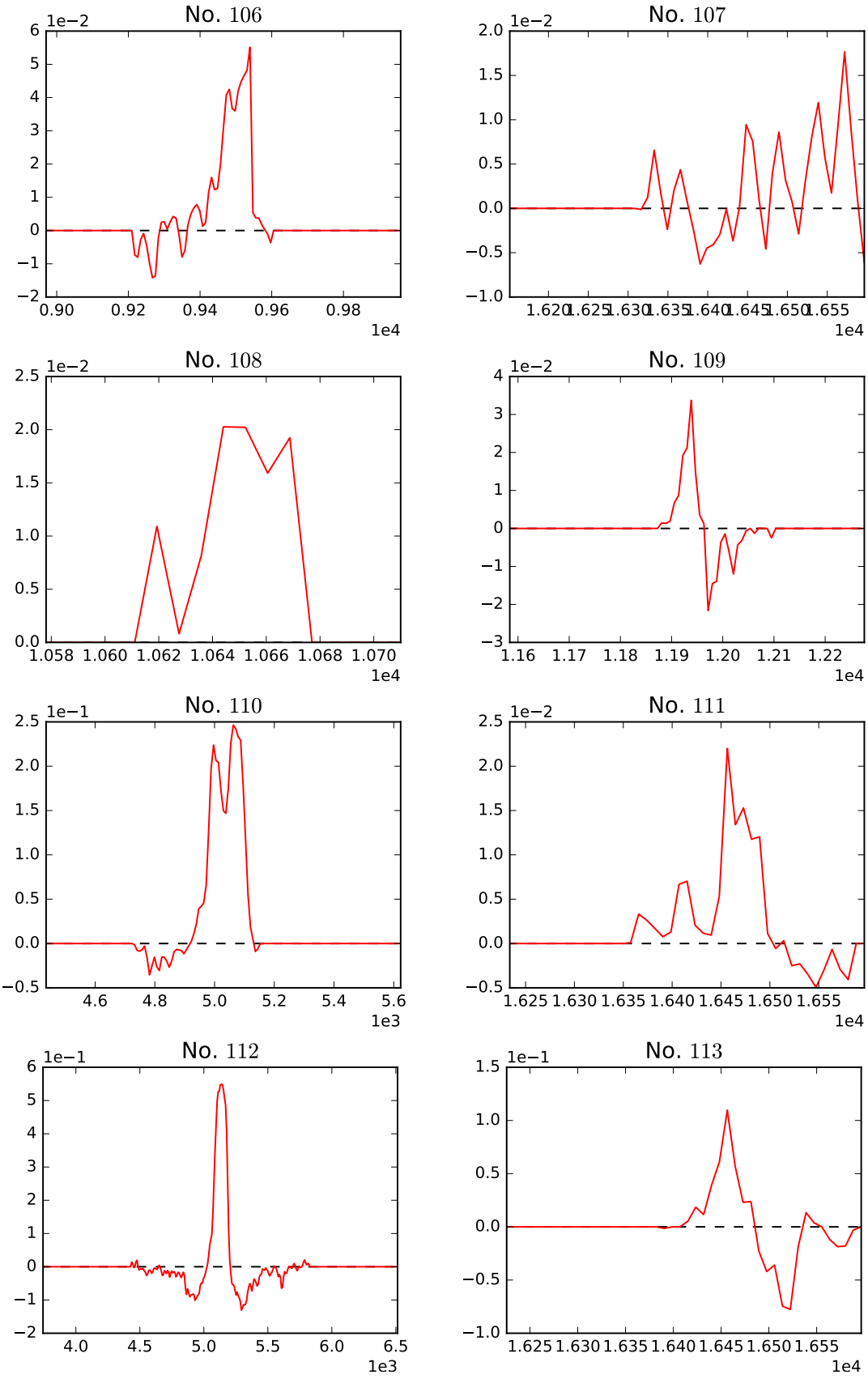

No. 100
1e−1
1.0
0.5
0.0
−0.5
−1.0
−1.5
1.10 1.15 1.20 1.25 1.30 1.35 1.40
1e4

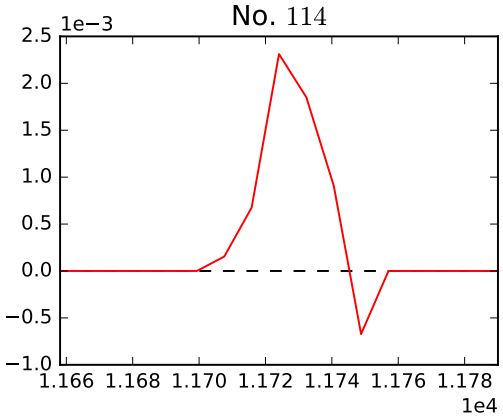

No. 101
1e−2
4
3
2
1
0
−1
−2
−3
−4
6.2 6.3 6.4 6.5 6.6 6.7 6.8 6.9 7.0
1e3

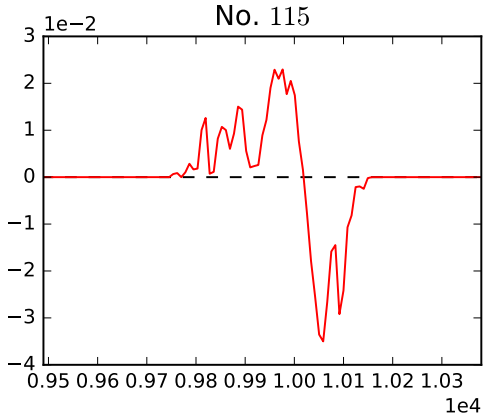

No. 102
1e−2
4
3
2
1
0
−1
−2
−3
1.20 1.22 1.24 1.26 1.28 1.30 1.32 1.34
1e4

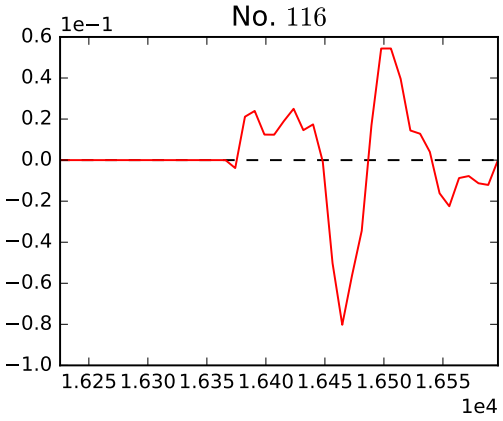

No. 103
1e−3
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.190 1.192 1.194 1.196 1.198 1.200 1.202 1.204
1e4

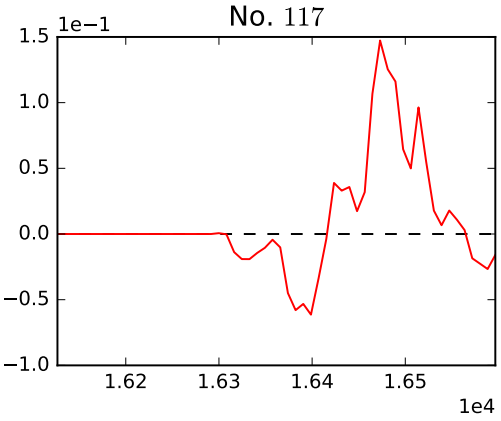

No. 105
1e−2
6
5
4
3
2
1
0
−1
−2
−3
1.20 1.25 1.30 1.35
1e4

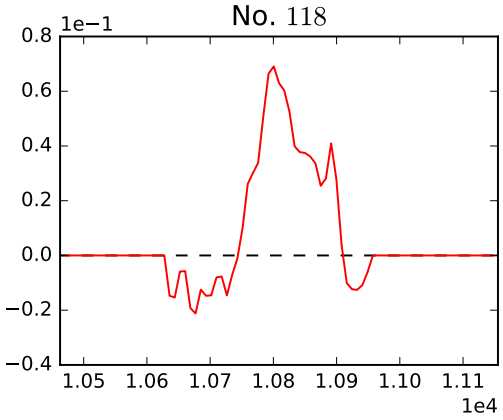

No. 106
1e−2
1e4
No. 107
1e−2
1e4
No. 108
1e−2
1e4
No. 109
1e−2
1e4
No. 110
1e−1
1e3
No. 111
1e−2
1e4
No. 112
1e−1
1e3
No. 113
1e−1
1e4

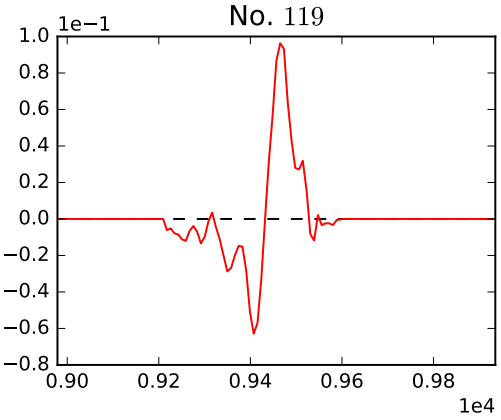

No. 114
1e−3
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
1.166 1.168 1.170 1.172 1.174 1.176 1.178
1e4

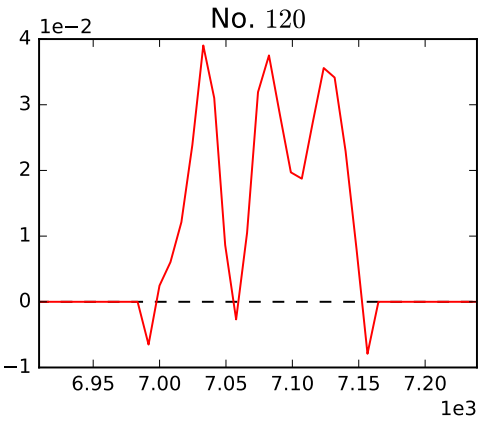

No. 115
1e−2
3
2
1
0
−1
−2
−3
−4
0.95 0.96 0.97 0.98 0.99 1.00 1.01 1.02 1.03
1e4

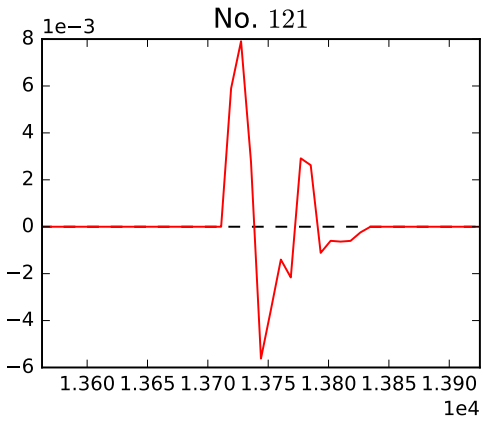

No. 116
1e−1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
1.625 1.630 1.635 1.640 1.645 1.650 1.655
1e4

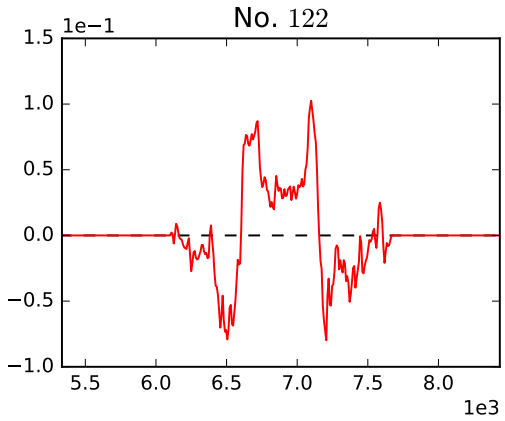

No. 117
1e−1
1.5
1.0
0.5
0.0
−0.5
−1.0
1.62 1.63 1.64 1.65
1e4

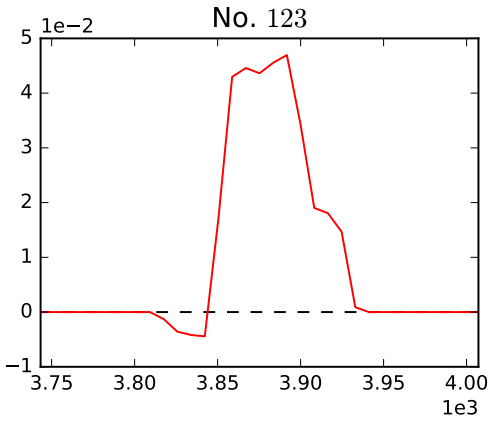

No. 118
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
1.05 1.06 1.07 1.08 1.09 1.10 1.11
1e4

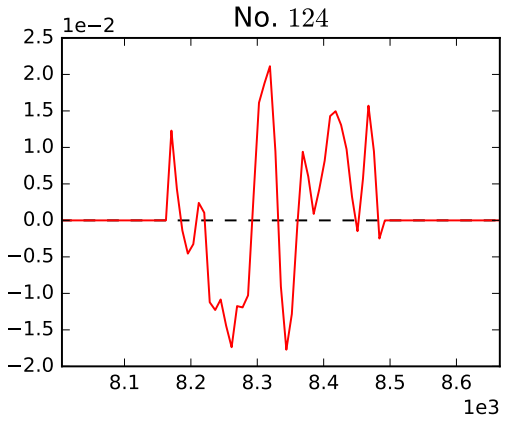

No. 119
1e−1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.90 0.92 0.94 0.96 0.98
1e4

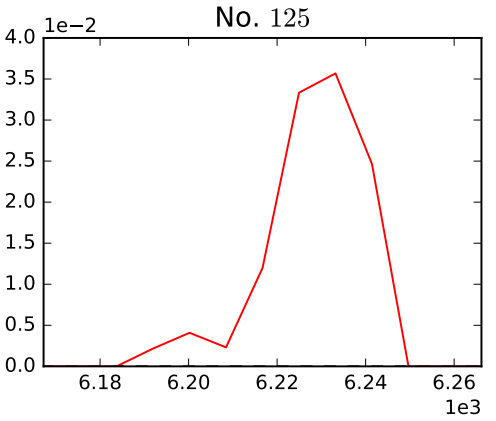

No. 120
1e−2
4
3
2
1
0
−1
6.95 7.00 7.05 7.10 7.15 7.20
1e3

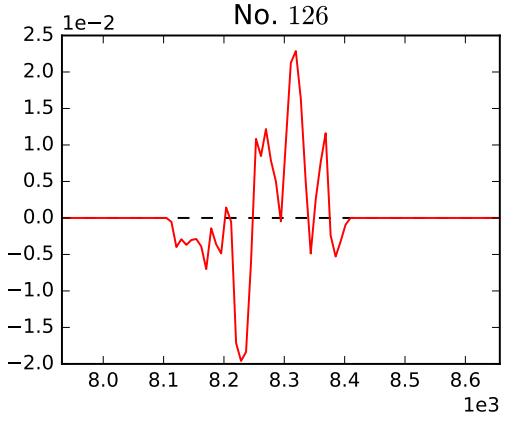

No. 121
1e−3
8
6
4
2
0
−2
−4
−6
1.360 1.365 1.370 1.375 1.380 1.385 1.390
1e4

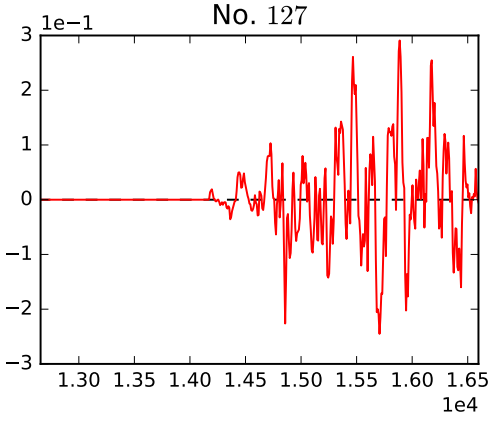

No. 122
1e−1
1.5
1.0
0.5
0.0
−0.5
−1.0
5.5
6.0
6.5
7.0
7.5
8.0
1e3

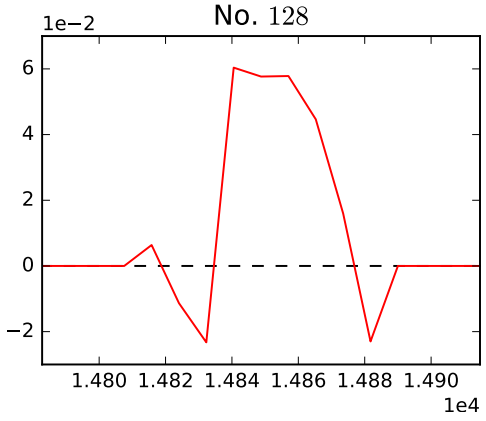

No. 123
1e−2
5
4
3
2
1
0
−1
3.75
3.80
3.85
3.90
3.95
4.00
1e3

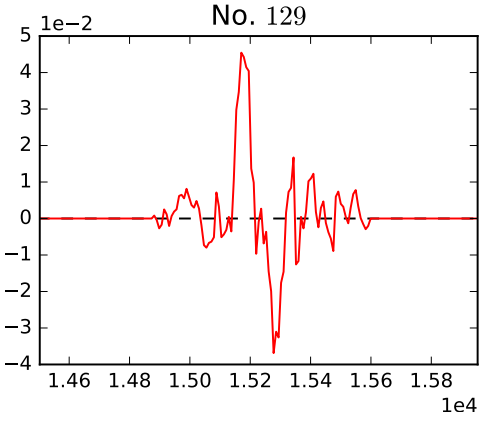

No. 124
1e−2
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
8.1
8.2
8.3
8.4
8.5
8.6
1e3

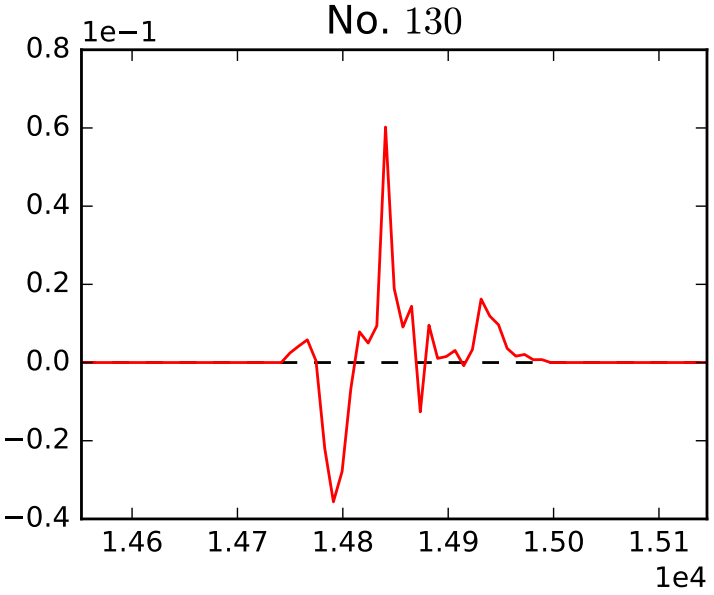

No. 125
1e−2
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
6.18
6.20
6.22
6.24
6.26
1e3

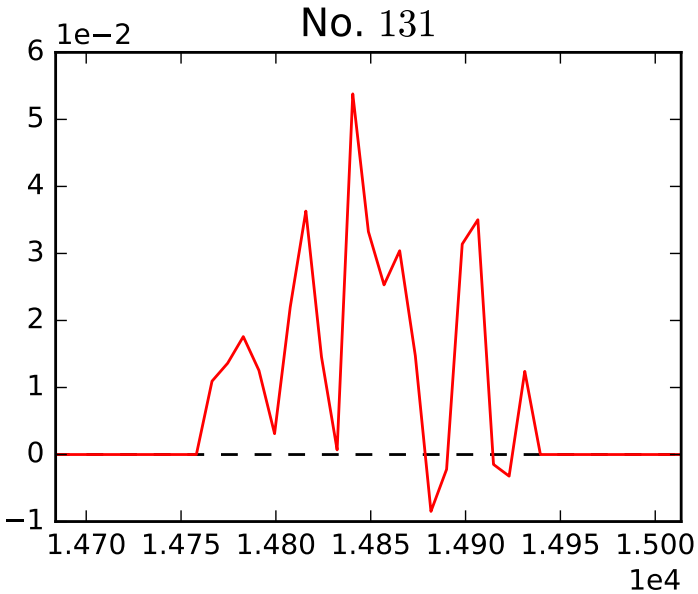

No. 126
1e−2
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
8.0
8.1
8.2
8.3
8.4
8.5
8.6
1e3


No. 127
1e−1
3
2
1
0
−1
−2
−3
1.30
1.35
1.40
1.45
1.50
1.55
1.60
1.65
1e4


No. 128
1e−2
6
4
2
0
−2
1.480
1.482
1.484
1.486
1.488
1.490
1e4


No. 129
1e−2
5
4
3
2
1
0
−1
−2
−3
−4
1.46
1.48
1.50
1.52
1.54
1.56
1.58
1e4


No. 130
1e−1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
1.46
1.47
1.48
1.49
1.50
1.51
1e4


No. 131
1e−2
6
5
4
3
2
1
0
−1
1.470
1.475
1.480
1.485
1.490
1.495
1.500
1e4

Appendix D

SoFiA parameter settings

- **Parameter activation**
  - steps.doSCfind
  - steps.doMerge
  - steps.doReliability
  - steps.doParameterise
  - steps.doWriteMask
  - steps.doWriteCat
  - steps.doMom0
  - steps.doMom1
- **Import data cube**
  - import.inFile: spectral cube
- **S + C Finder**
  - SCfind.threshold = 3.5
  - SCfind.sizeFilter = 0.0
  - SCfind.maskScaleXY = 2.0
  - SCfind.maskScaleZ = 2.0
  - SCfind.edgeMode = constant
  - SCfind.rmsMode = negative
  - SCfind.kernelUnit = pixel
  - SCfind.verbose = true
- **Merging detected voxels**
  - merge.radiusX = 3
  - merge.radiusY = 3
  - merge.radiusZ = 3
  - merge.minSizeX = 5
  - merge.minSizeY = 5
  - merge.minSizeZ = 5
  - merge.positivity = false
- **Parametrisation**
  - parameters.getUncertainties = false
  - parameters.fitBusyFunction = false
  - parameters.optimiseMask = true
  - parameters.dilateMask = false
  - parameters.dilateThreshold = 0.02
  - parameters.dilatePixMax = 10
  - parameters.dilateChan = 1
- **Output module**
  - writeCat.writeASCII = true
  - writeCat.parameters = ['id', 'name', 'ra', 'dec', 'vopt', 'snr_int', 'f_peak', 'f_int', 'w20', 'w50', 'rel']
- **Merging detected voxels**
  - merge.radiusX = 3
  - merge.radiusY = 3
  - merge.radiusZ = 3
  - merge.minSizeX = 5
  - merge.minSizeY = 5
  - merge.minSizeZ = 5
  - merge.positivity = false
- **Reliability calculation**
  - reliability.parSpace = ['n_pix','snr_sum','snr_max']
  - reliability.logPars = [1,1,1]
  - reliability.autoKernel = true
  - reliability.scaleKernel = 0.5
  - reliability.usecov = true
  - reliability.negPerBin = 5.0
  - reliability.skellamTol = -0.2
  - reliability.kernel = [0.15,0.05,0.1]
  - reliability.fMin = 10.0
  - reliability.threshold = 0.90
  - reliability.makePlot = true

D.1 SOFIA parameters fine-tuning

The parameters settings used in this project are based on the recommended values from the SOFIA user manual, however, we also conducted further fine-tuning of parameters so to suit our data cube. After running the pipeline for more than 20 times on a subcube while varying certain parameters, we found optimal parameters for our data cube. We ran the pipeline on the entire cube and obtained about 40% completeness and 55% reliability, we then selected seven parameters (i.e. SCfind.threshold, merge.radiusR = [merge.radiusX, merge.radiusY, merge.radiusZ], merge.minSize=[merge.minSizeX , merge.minSizeY and merge.minSizeZ]) and we varied them slightly and compared the results with both Visual and SA catalogues. A sample of these results is presented in the table below. It must be noted that the results below are inclusive of duplicates. The best results are given by Run 2.

TABLE D.1: Total number of SOFIA detections as a function of six parameters.

Run	**Changed parameters**	**SOFIA detections**	**Visual crossmatches**	**SA crossmatches**
Run 1	SCfind.threshold = 4.0 merge.radiusR = 3.0 merge.minSize = 5.0	119	89	83
Run 2	SCfind.threshold = 3.5 merge.radiusR = 3.0 merge.minSizeR= 5.0	137	97	102
Run 3	SCfind.threshold = 3.5 merge.radiusR = 3.0 merge.minSizeR = 4.0	120	88	80
Run 4	SCfind.threshold = 3.5 merge.radiusR = 2.0 merge.minSizeR = 5.0	107	83	92
Run 5	SCfind.threshold = 3.0 merge.radiusR = 3.0 merge.minSizeR = 5.0	115	92	84